



STIC Search Report

Biotech-Chem Library

STIC Database Tracking Number: 131674

To: Sarvamangala Devi
Location: REM 3C18
Art Unit: 1645
Friday, September 03, 2004

Case Serial Number: 09/743674

From: Beverly Shears
Location: Remsen Bldg.
RM 1A54

Phone: 571-272-2528

beverly.shears@uspto.gov

Search Notes

Shears, Beverly

131674

From: Devi, Sarvamangala
Sent: Wednesday, September 01, 2004 10:56 AM
To: Shears, Beverly
Subject: 09/743,674

Beverly:

Would you please perform a sequence and an interference search for SEQ ID NO: 2 in application 09/743,674?

Thanx.

S. DEVI, Ph.D.
AU 1645
Rems - 3C18



Date completed: 09-03-04
Searcher: Beverly e2528
Terminal time: 23
Elapsed time: _____
CPU time: _____
Total time: 23
Number of Searches: _____
Number of Databases: 1

Search Site

____ STIC
____ CM-1
____ Pre-S

Type of Search

____ N.A. Sequence
____ A.A. Sequence
____ Structure
____ Bibliographic

Vendors

____ IG
____ STN
____ Dialog
____ APS
____ Geninfo
____ SDC
____ DARC/Questel
____ Other CGN

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 2, 2004, 19:05:57 ; Search time 135 Seconds
(without alignments)
1841.791 Million cell updates/sec

Title: US-09-743-674-2
Perfect score: 4404
Sequence: 1 MPAGRLPRCPMTKFTDCT.....EEAEGDVLKRAALAQELGI 880

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A Geneseq 29Jan04:*
1: Geneseqp1980s:*
2: Geneseqp1990s:*
3: Geneseqp2000s:*
4: Geneseqp2001s:*
5: Geneseqp2002s:*
6: Geneseqp2003as:*
7: Geneseqp2003Bs:*
8: Geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4404	100.0	880	3	AAY44638 N. mening
2	3434	78.0	801	6	ABU37449 Protein e
3	2391.5	54.3	503	6	ABU37647 Protein e
4	1122.5	25.5	244	6	ABP80077 N. gonorr
5	1122.5	25.5	244	6	ABP78566 N. gonorr
6	1025	23.3	275	6	ABP78509 N. gonorr
7	1025	23.3	275	6	ABP78509 N. gonorr
8	821.5	18.7	253	6	ABP78538 N. gonorr
9	620.5	14.1	132	6	ABP79950 N. gonorr
10	381.5	8.7	919	6	ABU38527 Protein e
11	358	8.1	1095	6	ABJ26247 Aspergill
12	319.5	7.3	2768	4	ABP68397 Drosophil
13	315.5	7.2	1786	2	AAW24790 P. falcip
14	315.5	7.2	1787	5	AAU96699 Plasmodiu
15	315	7.2	1000	6	ABJ25647 Aspergill
16	313	7.1	1822	2	AAJ27745 Extracell
17	289.5	6.6	1468	4	ABJ27745 Extracell
18	289	6.6	8991	6	ABU08487 S. pneumo
19	287.5	6.5	753	4	ABJ27745 Extracell
20	277	6.3	2748	4	ABJ27745 Extracell
21	276.5	6.3	2468	7	ABJ27745 Extracell
22	276.5	6.3	2468	7	ABJ27745 Extracell
23	276.5	6.3	2468	7	ABJ27745 Extracell
24	276.5	6.3	2468	7	ABJ27745 Extracell
25	276.5	6.3	2468	7	ABJ27745 Extracell

26	276.5	6.3	2519	4	ABG16636 Novel hum
27	274.5	6.2	1192	4	ABJ59642 Drosophil
28	269	6.1	2364	6	ABM04816 Rat micro
29	269	6.1	2459	7	ABM04816 Rat micro
30	269	6.1	2459	7	ABM04816 Rat micro
31	269	6.1	2459	7	ABM04816 Rat micro
32	269	6.1	2459	7	ABM04816 Rat micro
33	269	6.1	2459	7	ABM04816 Rat micro
34	267	6.1	2459	7	ABM04816 Rat micro
35	256.5	5.8	8805	3	ABJ57274 E. canis
36	253	5.7	710	4	ABJ57274 E. canis
37	252	5.7	837	6	ABJ57274 E. canis
38	252	5.7	864	4	ABJ57274 E. canis
39	252	5.7	864	4	ABJ57274 E. canis
40	251	5.7	864	4	ABJ57274 E. canis
41	247	5.6	842	4	ABJ57274 E. canis
42	246	5.6	795	4	ABJ57274 E. canis
43	244	5.5	1616	6	ABJ57274 E. canis
44	244	5.5	1616	6	ABJ57274 E. canis
45	243.5	5.5	1558	3	ABJ57274 E. canis

ALIGNMENTS

RESULT 1
AAY44638
ID AAY44638 standard; protein; 880 AA.
XX AC
XX AAY44638;
DT 18-APR-2000 (first entry)
DE N. meningitidis T-cell stimulating protein A (TSPA).
XX T-cell stimulating protein A; TSPA; CD4+ T-cell; stimulant; meningitis;
KW antibacterial; anti-inflammatory; vaccine; neisserial disease;
KW gonorrhoea; septicaemia; septic arthritis; peivic inflammatory disease;
KW meningococcal; gonococcal.
XX Neisseria meningitidis.
XX WO200003003-A2.
XX 20-JAN-2000.
PD
XX 09-JUL-1999; 99WO-GB002205.
XX 10-JUL-1999; 98GB-00014902.
XX (UYNO-) UNIV NOTTINGHAM.
PI Ala'aldeen D, Todd I;
DR WPI; 2000-147612/13.
XX N-PSDB; AAZ49702.
PT Generation of cell lines and clones specific to a particular protein for
screening antigenic peptides which are used as vaccines in treating
meningococcal, gonococcal infections.
XX Claim 59; Page 42-45; 51pp; English.
CC The present sequence is N. meningitidis (strain SD, serogroup B
(B:15:P1.16)) T-cell stimulating protein A (TSPA). TSPA is a CD4+ T-cell
stimulant. It can be produced recombinantly using lambda ZapII phage
library comprising the DNA encoding TSPA. TSPA has antibacterial and anti
inflammatory activity and can be used in vaccine formulations against
neisserial diseases like meningitis, gonorrhoea, septicaemia, septic
arthritis and pelvic inflammatory diseases. T-cell lines and clones
specific to neisserial proteins can be generated for screening
meningococcal or gonococcal genomic phase display libraries to identify
peptides which stimulate T-cell lines and clones

23-OCT-2003. (revised)
19-JUN-2003 (first entry)
Protein encoded by Prokaryotic essential gene #22976.
Antisense; prokaryotic essential gene; cell proliferation; drug design.
Neisseria gonorrhoeae.
W0200277183-A2.
03-OCT-2002.
21-MAR-2002; 2002WO-US009107.
21-MAR-2001; 2001US-00815242.
06-SEP-2001; 2001US-00948993.
25-OCT-2001; 2001US-0342923P.
08-FEB-2002; 2002US-00072851.
06-MAR-2002; 2002US-0362699P.
(ELIT-) ELITRA PHARM INC.
Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
WPI; 2003-029926/02.
N-PSDB; ACA41319.
New antisense nucleic acids, useful for identifying proteins or screening
for homologous nucleic acids required for cellular proliferation to
isolate candidate molecules for rational drug discovery programs.
Claim 25; SEQ ID NO 63373; 1766pp; English.
The invention relates to an isolated nucleic acid comprising any one of
the 6213 antisense sequences given in the specification where expression
of the nucleic acid inhibits proliferation of a cell. Also included are:
(1) a vector comprising a promoter operably linked to the nucleic acid
encoding a polypeptide whose expression is inhibited by the antisense
nucleic acid; (2) a host cell containing the vector; (3) an isolated
polypeptide or its fragment whose expression is inhibited by the
antisense nucleic acid; (4) an antibody capable of specifically binding
the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
proliferation or the activity of a gene in an operon required for
proliferation; (7) identifying a compound that influences the activity of
the gene product or that has an activity against a biological pathway
required for proliferation, or that inhibits cellular proliferation; (8)
identifying a gene required for cellular proliferation or the biological
pathway in which a proliferation-required gene or its gene product lies
or a gene on which the test compound that inhibits proliferation of an
organism acts; (9) manufacturing an antibiotic; (10) profiling a
compound's activity; (11) a culture comprising strains in which the gene
product is overexpressed or underexpressed; (12) determining the extent
to which each of the strains is present in a culture or collection of
strains; or (13) identifying the target of a compound that inhibits the
proliferation of an organism. The antisense nucleic acids are useful for
identifying proteins or screening for homologous nucleic acids required
for cellular proliferation to isolate candidate molecules for rational
drug discovery programs, or for screening homologous nucleic acids
required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
K. pneumoniae or *P. aeruginosa*. The present sequence is encoded by one of
the target prokaryotic essential genes. Note: The sequence data for this
patent did not form part of the printed specification, but was obtained
in electronic format directly from WIPO at
ftp.wipo.int/pub/published_pct_sequences. (Updated on 23-OCT-2003 to
standardise OS field)
XX Sequence 801 AA;
XX
Query Match 78.0%; Score 3434; DB 6; Length 801;
Best Local Similarity 77.9%; Pred. No. 1.1e-196;
XX

RESULT 2
ABU37449
ID ABU37449 standard; protein; 801 AA.
XX
AC ABU37449;

RESULT 3
ABU37647
ID ABU
XX
AC ABU
XX

Db 1 NTAASETAESAQAASALDPTDGTGNAVSEPVQVSAEBETESGLFDGLFGGSYTL 60
QY 375 LLAGGGAALIALLLLLLAQSKARRTEESVPEEPDLDDAADDGIIITFAEVETPATPE 434
Db 61 LLAGGGAALIALLLLLLAQSKARRTEESVPEEPDLDDAADDGIIITFAEVETPATPE 120
QY 435 PAPKNDVNDTIALDGESEELSAKQTFDVTETDPSNRIDLDLDFSLAAQNGILSGALTQD 494
Db 121 PAPKNDVNDTIALDGESEELSAKQTFDVTETDPSNRIDLDLDFSLAAQNGILSGALTQD 180
QY 495 EETOKRADADWNALESVDVPEETPNPVEIVIDTPESVAQTAENKPKETVDTDFS 554
Db 181 EETOKRADADWNALESVDVPEETPNPVEIVIDTPESVAQTAENKPKETVDTDFS 240
QY 555 DNLPSNNHIGTEETASAKPSPGLAGFLKASSPETTILEKTVAEVQTPPELHDFLKVYET 614
Db 241 DNLPSNNHIGTEETASAKPSPGLAGFLKASSPETTILEKTVAEVQTPPELHDFLKVYET 300
QY 615 DAVAEAPETPDFNAADDLSALLQPAEAPSVENITETVAETPDFNATADDLSALLQPS 674
Db 301 DAVAEAPETPDFNAADDLSALLQPAEAPSVENITETVAETPDFNATADDLSALLQPS 360
QY 675 EVPAVEENAAEIVADDLSALLQPAEAPAEVENVETVAETSDFTHTAADDLSALLQPAEVP 734
Db 361 EVPAVEENAAEIVADDLSALLQPAEAPAEVENVETVAETSDFTHTAADDLSALLQPS 420
QY 735 AVENVETVAEIPDFNATADDLSALLQPSVPAVEENAAEITLETPTSNTSEADALPDF 794
Db 421 AVENAAETV-----ADDLSALLQPAEAPAEVENVETVAEIPDFNATADDLSALLQPS 471
QY 795 LKDGSEETVDWSIYLSEENIPNNADTSF 822
Db 472 LKDGSEETVDWSIYLSEENIPNNADTSF 499

RESULT 4
ABP80077
XX ID ABP80077 standard; protein; 244 AA.
XX AC ABP80077;
XX DT 07-MAR-2003 (first entry)
XX DE N. gonorrhoeae amino acid sequence SEQ ID 6684.
XX KW Antibacterial; infection; vaccine; gene therapy.
XX OS Neisseria gonorrhoeae.
XX PN WO200279243-A2.
XX PD 10-OCT-2002.
XX PF 12-FEB-2002; 2002WO-IB002069.
XX PR 12-FEB-2001; 2001GB-00003424.
XX PA (CHIR-) CHIRON SPA.
XX PI Fontana MR, Pizza M, Masignani V, Monaci E;
XX DR WPI; 2003-058415/05.
XX DR N-PSDB; ABZ41047.
XX PS New protein from Neisseria gonorrhoeae, useful for the manufacture of a
XX PT medicament for treating or preventing N. gonorrhoeae infection.
XX CC Disclosure; Page 674; 815pp; English.
XX CC The present invention relates to proteins from Neisseria gonorrhoeae.
XX CC Also disclosed are the nucleic acid molecules encoding the proteins and
XX CC antibodies that specifically bind to the proteins. The composition

comprising the protein, nucleic acid or antibody is useful for the
manufacture of a medicament for treating or preventing N. gonorrhoeae
infection, this may be in the form of a vaccine or gene therapy.
Sequences given in records ABP76736-ABP81046 represent nucleic acid
molecules of the invention
SQ Sequence 244 AA;
Query Match 25.5%; Score 1122.5; DB 6; Length 244;
Best Local Similarity 64.6%; Pred. No. 2.9e-59;
Matches 237; Conservative 2; Mismatches 5; Indels 123; Gaps 1;
QY 514 VYBETPNPNPVEIVIDTPESVAQTAENKPKETVDTDFSNNHIGTEETASAKP 573
Db 1 VYBETPNPNPVEIVIDTPESVAQTAENKPKETVDTDFSNNHIGTEETASAKP 60
QY 574 ASPGSLAGFLKASSPETTILEKTVAEVQTPPELHDFLKVYETDAVAETAPETPDFNAADD 633
Db 61 AAPGSLAGFLKASSPETTILEKTVAEVQTPPELHDFLKVYETGAETAPETPDFNAADD 120
QY 634 LSALLQPAEAPSVENITETVAETPDFNATADDLSALLQPSVPAVEENAAEIVADDLSA 693
Db 121 LSALLQPAEA----- 130
QY 694 LLQPAEAPAEVENVETVAETSDFTHTAADDLSALLQPAEVPVENVETKTVAEIPDFNAT 753
Db 131 ----- 130
QY 754 ADDLSALLQPSVPAVEENAAEITLETPTSNTSEADALPDFLKDGSEETVDWSIYLSEEN 813
Db 131 -----PAVEENAAEITLETPTSNTSEADALPDFLKDGSEETVDWSIYLSEEN 177
QY 814 IPNNADTSFPSESVGSDAPSEAKYDLAEMYLEIGORDAAETVQKLEAAGDVLKRAQA 873
Db 178 IPNNADTSFPSESVGSDAPSEAKYDLAEMYLEIGORDAAETVQKLEAAGDVLKRAQA 237
QY 874 LAQELGI 880
Db 238 LAQELGI 244
RESULT 5
ABP78566
XX ID ABP78566 standard; protein; 244 AA.
XX AC ABP78566;
XX DT 07-MAR-2003 (first entry)
XX DE N. gonorrhoeae amino acid sequence SEQ ID 3662.
XX KW Antibacterial; infection; vaccine; gene therapy.
XX OS Neisseria gonorrhoeae.
XX PN WO200279243-A2.
XX PD 10-OCT-2002.
XX PF 12-FEB-2002; 2002WO-IB002069.
XX PR 12-FEB-2001; 2001GB-00003424.
XX PA (CHIR-) CHIRON SPA.
XX PI Fontana MR, Pizza M, Masignani V, Monaci E;
XX DR WPI; 2003-058415/05.
XX DR N-PSDB; ABZ39536.
XX PS New protein from Neisseria gonorrhoeae, useful for the manufacture of a
XX PT medicament for treating or preventing N. gonorrhoeae infection.
XX CC

PS Disclosure; Page 453; 815pp; English.

XX The present invention relates to proteins from *Neisseria gonorrhoeae*.

CC Also disclosed are the nucleic acid molecules encoding the proteins and

CC antibodies that specifically bind to the proteins. The composition

CC comprising the protein, nucleic acid or antibody is useful for the

CC manufacture of a medicament for treating or preventing *N. gonorrhoeae*

CC infection, this may be in the form of a vaccine or gene therapy.

CC Sequences given in records ABP76736-ABP81046 represent nucleic acid

CC molecules of the invention

XX Sequence 244 AA;

SQ

Query Match 25.5%; Score 1122.5; DB 6; Length 244;

Best Local Similarity 64.6%; Pred. No. 2.9e-59;

Matches 237; Conservative 2; Mismatches 5; Indels 123; Gaps 1;

QY 514 VYEPETNPYNPVEIVIDTPEPSVAQTAENKPTVDTDFSDNLPNNHIGTEETASAKP 573

DB 1 VYEPETNPYNPVEIVIDTPEPSVAQTAENKPTVDTDFYNNLFSNNHIGTEETASAKP 60

QY 574 ASPGSLAGFLKASPETILEKTVAEVQTPPELHDFLKVYETDAVETAPETPDFAAAD 633

DB 61 AAPGSLAGFLKASPETILEKTVAEVQTPPELHDFLKVYETGAVETAPETPDFAAAD 120

QY 634 LSALLQPAEAPSVEENITETVETPDNPATADDLSALLQPSVEPAVEENAAEIVADDLSA 693

DB 121 LSALLQPAEA----- 130

QY 694 LLQPAEAPAVEENVETVETVETSDPHTAADDLSALLQPAEPAVEENVTKTVAEIPDPNAT 753

DB 131 ----- 130

QY 754 ADDLSALLQPSVEPAVEENAAEITLTPDSNTSEADALPDFLKGDEETVDSIYLSEEN 813

DB 131 -----PAVEENAAEITLTPDSNTSEADALPDFLKGDEETVDSIYLSEEN 177

QY 814 IPNADTSPFSEVSGSDAPSEAKYDLAEMYLEIGDRDAAAEYVQKLEEGDVLKRAQA 873

DB 178 IPNADTSPFSEVSGSDAPSEAKYDLAEMYLEIGDRDAAAEYVQKLEEGDVLKRAQA 237

QY 874 LAQELGI 880

DB 238 LAQELGI 244

RESULT 6

ABP78509

ID ABP78509 standard; protein; 275 AA.

AC ABP78509;

XX

DT 07-MAR-2003 (first entry)

DE N. gonorrhoeae amino acid sequence SEQ ID 3548.

XX Antibacterial; infection; vaccine; gene therapy.

XX *Neisseria gonorrhoeae*.

OS

XX WO200279243-A2.

PN

PD 10-OCT-2002.

XX

PF 12-FEB-2002; 2002WO-IB002069.

XX

PR 12-FEB-2001; 2001GB-00003424.

XX

PA (CHIR-) CHIRON SPA.

PI Fontana MR, Pizza M, Massignani V, Monaci E;

XX WPI; 2003-058415/05.

DR

DR N-PSDB; ABZ39479.

XX

PT New protein from *Neisseria gonorrhoeae*, useful for the manufacture of a

PT medicament for treating or preventing *N. gonorrhoeae* infection.

XX

PS Disclosure; Page 447; 815pp; English.

XX

CC The present invention relates to proteins from *Neisseria gonorrhoeae*.

CC Also disclosed are the nucleic acid molecules encoding the proteins and

CC antibodies that specifically bind to the proteins. The composition

CC comprising the protein, nucleic acid or antibody is useful for the

CC manufacture of a medicament for treating or preventing *N. gonorrhoeae*

CC infection, this may be in the form of a vaccine or gene therapy.

CC Sequences given in records ABP76736-ABP81046 represent nucleic acid

CC molecules of the invention

XX Sequence 275 AA;

SQ

Query Match 23.3%; Score 1025; DB 6; Length 275;

Best Local Similarity 83.4%; Pred. No. 2.2e-53;

Matches 216; Conservative 3; Mismatches 6; Indels 34; Gaps 3;

QY 107 LGDKAVIAVSSEQAVRDPVLVFRIGAGAQVREYTAIILDPVGYSPKTKSALSDGKTHRKTA 166

DB 1 LDGKAVIAVSQAQAVRDPVLVFRIGAGAQVREYTAIILDPVGYSPKTKSALSDGKTHRKTA 60

QY 167 PTAESQENQNAKALRKTDKKDSANAAVKPAINGKTHTVRKGETVKQIAAAAIRPKHLTLEQ 226

DB 61 PKAESQENQNAKALRKTDKKDSANAAVKPAINGKTHTVRKGETVKQIAAAAIRPKHLTLEQ 120

QY 227 VADALLKANPNVSAHGLRAGSVLHPIPNLR-----IKAEQPKQTAKPAAETASMP 278

DB 121 VADVLLKANPNVSAHGLRAGSVLHPIPNLR-----IKAEQPKQTAKPAAETASMP 180

QY 279 SEPSKQAT-----VEKPVKEPAKVAAPAEKAEKPAVRP----- 313

DB 181 SEPSKQATVEKPIEKPVKEPAKVAAPAEKAEKPAVRP----- 313

QY 314 -PVPAANTAASETAASAP 331

DB 241 GPVPAANTAASETAASAP 259

RESULT 7

ABP79684

ID ABP79684 standard; protein; 275 AA.

XX

AC ABP79684;

XX

DT 07-MAR-2003 (first entry)

DE N. gonorrhoeae amino acid sequence SEQ ID 5898.

XX Antibacterial; infection; vaccine; gene therapy.

XX *Neisseria gonorrhoeae*.

OS

XX WO200279243-A2.

PN

PD 10-OCT-2002.

XX

PF 12-FEB-2002; 2002WO-IB002069.

XX

PR 12-FEB-2001; 2001GB-00003424.

XX

PA (CHIR-) CHIRON SPA.

PI Fontana MR, Pizza M, Massignani V, Monaci E;

XX WPI; 2003-058415/05.

DR N-PSDB; ABZ40654.

XX

PT New protein from *Neisseria gonorrhoeae*, useful for the manufacture of a

PT medicament for treating or preventing N. gonorrhoea infection.

XX Disclosure; Page 619; 815pp; English.

XX The present invention relates to proteins from Neisseria gonorrhoeae.
 CC Also disclosed are the nucleic acid molecules encoding the proteins and
 CC antibodies that specifically bind to the proteins. The composition
 CC comprising the protein, nucleic acid or antibody is useful for the
 CC manufacture of a medicament for treating or preventing N. gonorrhoea
 CC infection, this may be in the form of a vaccine or gene therapy.
 CC Sequences given in records ABP76736-ABP81046 represent nucleic acid
 CC molecules of the invention

XX Sequence 275 AA;

Query Match 23.3%; Score 1025; DB 6; Length 275;
 Best Local Similarity 83.4%; Pred. No. 2.2e-53;
 Matches 216; Conservative 3; Mismatches 6; Indels 34; Gaps 3;

QY 107 LGDKAVTAVSFEQAVRDPVLVFRIGAGAOVREYTAILDPVGYSPKTKSALSDGKTHRKA 166
 DB 1 LDGKAVTAVSQAQVRDPVLVFRIGAGAOVREYTAILDPVGYSPKTKSALSDGKTHRKA 60
 QY 167 PTASQENQNAKLRKTDKDSANAAVKPAYNGKTHVVRKGETVKQIAAARPKHLTLEQ 226
 DB 61 PKAESQENQNAKLRKTDKDSANSAVKPAHNGKTHVVRKGETLQIAAARPKHLTLEQ 120
 QY 227 VADALLKANPNVSAHGRURAGSVLHLPNLR-----IKAEQPKQTAETASMP 278
 DB 121 VADVLKANPNVSAHGRURAGSVLHLPNLR-----IKAEQPKQTAETASMP 180
 QY 279 SEPSKQAT-----VKPKVEKPEAKVAAPKAEPKAVRPE----- 313
 DB 181 SEPSKQATVEKPEKVEKPEKVAAPKAEPKAVRPEKPAVSETPASATERQP 240

QY 314 -PVPAAANTAASETAESAP 331
 DB 241 GPVPAANTAASETAESAP 259

RESULT 8
 ABP78538
 ID ABP78538 standard; protein; 253 AA.

XX AC ABP78538;
 XX DT 07-MAR-2003 (first entry)
 XX DE N. gonorrhoeae amino acid sequence SEQ ID 3606.
 XX KW Antibacterial; infection; vaccine; gene therapy.

XX Neisseria gonorrhoeae.

XX WO200279243-A2.

XX PD 10-OCT-2002.

XX PF 12-FEB-2002; 2002WO-1B002069.

XX PR 12-FEB-2001; 2001GB-00003424.

XX PA (CHIR-) CHIRON SPA.

XX Fontana MR, Pizza M, Massignani V, Monaci E;

XX WPI; 2003-058415/05.

XX N-PSDB; ABZ39508.

XX New protein from Neisseria gonorrhoeae, useful for the manufacture of a

PT medicament for treating or preventing N. gonorrhoea infection.

XX Disclosure; Page 450; 815pp; English.

XX The present invention relates to proteins from Neisseria gonorrhoeae.
 CC Also disclosed are the nucleic acid molecules encoding the proteins and
 CC antibodies that specifically bind to the proteins. The composition
 CC comprising the protein, nucleic acid or antibody is useful for the
 CC manufacture of a medicament for treating or preventing N. gonorrhoea
 CC infection, this may be in the form of a vaccine or gene therapy.
 CC Sequences given in records ABP76736-ABP81046 represent nucleic acid
 CC molecules of the invention

XX Sequence 253 AA;

Query Match 18.7%; Score 821.5; DB 6; Length 253;
 Best Local Similarity 66.8%; Pred. No. 2.9e-41;
 Matches 181; Conservative 11; Mismatches 40; Indels 39; Gaps 5;

QY 255 LNRKABQPKPQTAQKPAETASMPSEPSKQATVEKPEKVAAPKAEPKAVRPEP 314
 DB 1 LKNLKQKLP-----KQKRNRPD-PNPNPQCPLRLRRQNVANPDL 42
 QY 315 VPAANTAASETAESAPQEAASAIIDTPTDETGNVS-----BPVEQVS 358
 DB 43 YPLQILPHRKPLPNPPQEAASAIIDTPTDETGNVS-----BPVEQVS 102
 QY 359 ABEETES-----GLFGGSYTLILAGGGAALIA-LLLLLRLAQSKRARTTESVPPEPDLD 413
 DB 103 ABEETESGLFDGLFGGSYTLILAGGGAALIA-LLLLLRLAQSKRARTTESVPPEPDLD 162
 QY 414 DAADDGIEITFAEVETPATPEAPKNDVNDTLALDGESEBELSAKQTFDVTDTPSNRID 473
 DB 163 DAADDGIKITFAEVETPATPEAPKNDVNDTLALDGESEBELSAKQTFDVTDTPSNRID 222
 QY 474 LDFDLSAAQNGILSGALTQDEETOKRADAD 504
 DB 223 LDFDLSAAQNGILSGALTQDEETOKRADAD 253

RESULT 9
 ABP79950
 ID ABP79950 standard; protein; 132 AA.

XX AC ABP79950;
 XX DT 07-MAR-2003 (first entry)
 XX DE N. gonorrhoeae amino acid sequence SEQ ID 6430.
 XX KW Antibacterial; infection; vaccine; gene therapy.

XX Neisseria gonorrhoeae.

XX WO200279243-A2.

XX PD 10-OCT-2002.

XX PF 12-FEB-2002; 2002WO-1B002069.

XX PR 12-FEB-2001; 2001GB-00003424.

XX PA (CHIR-) CHIRON SPA.

XX Fontana MR, Pizza M, Massignani V, Monaci E;

XX WPI; 2003-058415/05.

XX N-PSDB; ABZ40920.

XX New protein from Neisseria gonorrhoeae, useful for the manufacture of a

PT medicament for treating or preventing N. gonorrhoea infection.

XX Disclosure; Page 656; 815pp; English.

XX The present invention relates to proteins from Neisseria gonorrhoeae.

CC Also disclosed are the nucleic acid molecules encoding the proteins and

antigenic nucleic acid; (4) an antibody capable of specifically binding the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular proliferation or the activity of a gene in an operon required for proliferation; (7) identifying a compound that influences the activity of the gene product or that has an activity against a biological pathway required for proliferation, or that inhibits cellular proliferation; (8) identifying a gene required for cellular proliferation or the biological pathway in which a proliferation-required gene or its gene product lies or a gene on which the test compound that inhibits proliferation of an organism acts; (9) manufacturing an antibiotic; (10) profiling a compound's activity; (11) a culture comprising strains in which the gene product is overexpressed or underexpressed, (12) determining the extent to which each of the strains is present in a culture or collection of strains; or (13) identifying the target of a compound that inhibits the proliferation of an organism. The antisense nucleic acids are useful for identifying proteins or screening for homologous nucleic acids required for cellular proliferation to isolate candidate molecules for rational drug discovery programs, or for screening homologous nucleic acids required for proliferation in cells other than *S. aureus*, *S. typhimurium*, *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of the target prokaryotic essential genes. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published/pct_sequences

582 LKLMVEYAEVMDREGFARQENELREIGGAQ-POVEQLKSRV-----PAMVAVAAVA 631
 535 PESVAQTAENKPTVDTDFSDNIPSNHHITETASAKP-ASPSGLAGFLKASSPETILE 593
 632 GLAGAKLAQDELDSFSLD-DLSLDDSGH-----AAKPDAAAGQDLDADFSLDDLLGGD 683
 594 KTVAEVQTPPE-ELHDFLKVVETDAVATAPETPDENAAADDLSALLQPAAPAPSEENITE 652
 684 DVQADLKSDGALDLDLTDSDLDLAASTPADKP-----VDDLDPLGLDFAR----- 728
 653 TVAETPDENATADLSALLQSEVPVAVENNAEIVADDLSALLQ---PARAPAVEENITE 709
 729 -LAETPS-QPRHDLGDFSLDLDAP--EDKLSQ---DFFLILNDEVPAAAPADNEFTLD 781
 710 TVAETSPHTAAD--DLSALLQPAEVPAVENVTVAETPDENATADLSALLQSEVP 767
 782 TEAAEPPALSLPDDFDLSLADDEPTPAPEKGEDEFAAQULDEVSAQULDELAS----- 833
 768 AVENNAEITLETDSNT---SEADALPDFLKDGHEETVDSIYLSSEENIPNNADTSPPS 824
 834 -----NLDEPKSATPSFAEDAASALDGD-----ADDDDFD 866
 825 ESVGSDAPSEAKYDLAEMYLIGDRDAAAEVTKVQLLEAEAGDVLKRAQA 873
 867 LSGADERAAT--KLDLARAYIDMGSEGDILDEVIL--AEGNDSQQA 911

RESULT 11
 ID ABJ26247
 XX AC ABJ26247
 XX DT 17-APR-2003 (first entry)
 XX DE Aspergillus fumigatus essential gene protein #905.
 XX KW Fungicide; cytostatic; essential gene; Aspergillus fumigatus; infection;
 XX KW cancer; contamination; biofilm; antibody; immune response.
 XX OS Aspergillus fumigatus.
 XX PN WO200286090-A2.
 XX PD 31-OCT-2002.
 XX PF 23-APR-2002; 2002WO-US013142.
 XX PR 23-APR-2001; 2001US-0285697P.
 XX PR 27-APR-2001; 2001US-0287066P.
 XX PR 05-JUN-2001; 2001US-0295890P.
 XX PR 09-JUL-2001; 2001US-0303899P.
 XX PR 31-AUG-2001; 2001US-0316362P.
 XX PA (ELIT-) ELITRA PHARM INC.
 XX PI Jiang B, Tishkoff D, Zamudio C, Eroshkin AM, Hu W, Lemieux SM;
 XX WPI; 2003-093124/08.

XX New purified or isolated nucleic acids of essential genes of Aspergillus
 XX fumigatus, useful for treating or preventing infections by A. fumigatus,
 XX or for treating a non-infectious disease in a subject e.g. cancer.
 XX Disclosure; Page: 175pp; English.
 XX The invention relates to novel purified or isolated nucleic acids of
 XX essential genes of Aspergillus fumigatus. The isolated nucleic acids of
 XX the invention are used to treat or prevent infections by a pathogenic
 XX organism such as A. fumigatus, to treat a non-infectious disease in a
 XX subject (e.g. cancer), to prevent or contain contamination of an object
 XX by A. fumigatus, or to prevent or inhibit formation on a surface of a
 XX biofilm comprising A. fumigatus. The polynucleotides are useful for

expressing recombinant protein for characterisation, screening or
 therapeutic use, as markers for host tissues in which the pathogenic
 organisms invade or reside, for comparing with the DNA sequence of A.
 fumigatus to identify duplicated genes or paralogues having the same or
 similar biochemical activity and/or function, for comparing with DNA
 sequences of other related or distant pathogenic organisms to identify
 potential orthologous essential or virulence genes, for selecting and
 making oligomers for attachment to a nucleic acid array for examination
 of expression patterns, for raising anti-protein antibodies, as an
 antigen to raise anti-DNA antibodies or to elicit another immune
 response, and for identifying polynucleotides encoding the other protein
 with which binding occurs or to identify inhibitors of the binding
 interaction. The polypeptides may be used to raise antibodies or to
 elicit immune response, as a reagent in assays designed to quantitatively
 determine levels of the protein in biological fluids, as a marker for
 host tissues in which pathogenic organism invade or reside, and to
 isolate correlative receptors or ligands in the case of virulence
 factors. This sequence represents a protein of one of the essential genes
 of Aspergillus fumigatus of the invention

Sequence 1095 AA;
 Query Match 8.1%; Score 358; DB 6; Length 1095;
 Best Local Similarity 22.1%; Pred. No. 9.5e-13;
 Matches 233; Conservative 140; Mismatches 412; Indels 270; Gaps 44;
 QY 38 NRQIKLIAASVAAAFQAHAGLGLGMLNQSLNLPFFSGSIIVTGEAKALGGGVTVSE 97
 DB 91 NNILTVGSSSIAAAPGAHA-----VEKETETPLENGASEKVEETAKEPGVQSTTTTE 144
 QY 98 KGLTAKVHKLGDKAVIA-----VSSFOVRD--PVLVFRIGAGAOVRE--YTALIDPVGYS 149
 DB 145 AVKDEVPEKTDSDSVVVEKDLHVESEPAEQVVKPLVSEPKSPDIHEDAETQPSANAAE 204
 QY 150 PKTKSALS-----DGKTHRTKTAFTAESQENQNAKALRKTDKOSANAAVKPAYNGK 200
 DB 205 PTNGTHSSEISGSQPDPAVTEKVEPVEEKAADLTTKVKANKPEEIPSTSLAEKPE 264
 QY 201 THTVRKGE-----TVKQIAAIRPKHL---TLEQVADALLKANPNVSAHGLRL 245
 DB 265 TKEVEKAEDQKLQDLIVPAIVPAIVEKTEAAKESEVPESTLEKSOEAAAPQ--TVTEVPLV 323
 QY 246 AGSVLHINLNRIKAEQ--PKQTA--KPKAETAS---MPSEP-----SKQATV----- 287
 DB 324 DSKPVLSAESTKAEKAPTMTTDEAPLAEKSSKTVPAEIAEKVQATIGDEPLSRN 383
 QY 288 ---EKPVKPK-EAKVAPEAKAEKPAVRPEPVAANTAASETAASASQAEAAASAI-DTP 342
 DB 384 NVVEEQPKQSVKANKAADAENVLEEP-----IKESAPBEIPEDSRNVAVDAP 430
 QY 343 TDETG-----NAVSEPVQVSARETESGLFGGSYTLILLAGGGAALLALLLLRL 392
 DB 431 VTESSTTEKVVTPADPAEKEPAKESAKEPIQEEIPEASEKEVAE-----TPV 479
 QY 393 AOSKRAARTEESV--PEREPDLDDAADDGI-----EITPA-EVETPATPEPAP 437
 DB 480 TESSSTTEKVEVTDALDAEKEPALEESTKEPIHBEQIETVVTAEVTAKEPEAQATEPVA 539
 QY 438 KNDVNDTLALDGESBE---ELSAKQTFDVT-----DTPSNRIDLDSDLSAAQNGI 486
 DB 540 NESINK--EVEATTEPTAGQSAKEPISNETAIRVTQADESTKELSTGF---AAADIT 594
 QY 487 LSGALTQDDET-----QKRAADWNAIESTD-----512
 DB 595 GKGAETHTEVSTAAGPAKEFVNDASAVKSAPTETVQETKDDSAQVVGKAEAPQIAEE 654
 QY 513 -SVYEPETNPVNPVEIVIDTPEP-----ESVAQTAENKPEVTD 550
 DB 655 AAVEETSAANPEAPTEESTTAEPAKEPGESEETPCQTLPAEESKEATAEAAKEBSTIE 714
 QY 551 TDFSDNL-----PSNNHIGTETASAKPSPGLAGFLKASSPETILEKTVAEVQTPPE 604
 DB 715 T-VSEPLEADVKPEAQDDPATETLTEDKPVQTSVAE--AAPKPESEVPVSEVSGTAKTS 771

Db 1988 DVATSTAPVAGGDIKDEQA 2008

RESULT 13

AAW24790
ID AAW24790 standard; protein; 1786 AA.

XX AAW24790;

XX 08-OCT-1997 (first entry)

XX P. falciparum liver stage antigen-3.

XX Plasmodium falciparum; pre-erythrocyte; liver stage antigen; serum;
KW prophylaxis; Thai strain; gene organisation; exon; intron; hydrophobic;
KW glycosyl-phosphatidylinositol membrane anchoring sequence; antibody;
KW vaccine; immunotherapy; malaria.

XX Plasmodium falciparum.

XX Key Location/Qualifiers

FT Region 223..278

FT /note= "repeat region 1"

FT Region 279..818

FT /note= "repeat region 2"

FT Region 1537..1576

FT /note= "repeat region 3"

XX WO9641877-A2.

XX 27-DEC-1996.

XX 12-JUN-1996; 96WO-FR000894.

XX 13-JUN-1995; 95FR-00007007.

XX (INSP) INST PASTEUR.

XX Drulhe P, Daubersies P;

XX WPI; 1997-065464/06.

XX N-PSDB; AAT78868.

XX Plasmodium falciparum poly:peptide(s) and related nucleic acids - derived
from the liver stage antigen-3, useful for malaria vaccine prodn. and
diagnosis.

XX Claim 1; Fig 2A-I; 69pp; French.

XX This sequence corresponds to a Plasmodium falciparum strain K1 pre-
erythrocytic liver stage antigen-3 (LSA-3) protein. The encoding gene
sequence was isolated by screening a P. falciparum strain T9/96 library
with serum from a missionary treated by prophylaxis (for strain T6/96 see
FR9101286). Of 20 clones isolated, clone 7295 was used to screen a
library generated from Thai strain K1. One clone contained a 6.85 kb
insert including the genomic sequence AAT78867. The gene comprises a 1.8
kb region encoding 3 major blocks of tetrapeptide repeats (especially the
amino acid sequence VEES, VEEN, VEEL, VAPS, VAPT, etc) and a 3'
hydrophobic region corresponding to a glycosyl-phosphatidyl- inositol
membrane anchoring sequence. The invention relates to new polypeptides of
at least 10 amino acids derived from the LSA-3 protein with the exception
of the peptides AAW24791-4. The LSA-3 peptides can be used to raise
antibodies and as vaccines for immunotherapy of malaria

XX Sequence 1786 AA;

SQ Query Match 7.2%; Score 315.5; DB 2; Length 1786;

Best Local Similarity 20.2%; Pred. No. 6.3e-10;

Matches 167; Conservative 147; Mismatches 331; Indels 183; Gaps 26;

XX 64 NTQSNLDPEFGSGITVGEAKALLGGGVTVSEKGLTAKVHLGDKAV---IATVSEQA 120

XX 242 NVEENVEENDGGSVASSVEES-----IASSVDESISSTIENVAPTVEEI 286

Qy	121	VRDPVLVFRIGAGQVREYTAILDPVGYSPKTKSALSDBGKTKHTAPTAE--QENQNAK	178
Db	287	VAPSVV-----ESVAPSVEESVEENVEES	310
Qy	179	ALRKTDKDSANAANKPAYNGKTHTVRKGE-----TVKQIAAAIRPKHLEQ--VADALL	232
Db	311	VAENVEESVAENVEESVAENVEESVAENVEEIVAPTVEEIVAP-----TVEEIVAPSVV	364
Qy	233	KA-NPNVSAHGRLRAGSVLHLPNLRKAKQPKPTAKPKAETAASMPSEKQATVSKPV	291
Db	365	ESVAPSVEESVE-----ENVEESVAENVEESVAENVEESVAENVEESV	415
Qy	292	EKPEAKVAPE-----AKAEKPAVRPEPVAANTAASETAESAPOQAAASAITPTDET	346
Db	416	AENVEEIVAPTVEEIVAPTVEEIVAPSVVSVAPSVEESVEENVEESVAENVEESVAENV	475
Qy	347	GNVSEPVQVSAEEETESGLFGGSYTLALLGGGAALIALLLLRQAQSKRARTTESVP	406
Db	476	EESVAENVEESVAENVEES-----VAENVEESVA	504
Qy	407	EE-EPDLDDAADGGIEITFAEVETPATPEPAPKNDVNDTLALDGESEEEESAKOTFDET	465
Db	505	ENVEESVAENVEEIVAPTVEEIVAPTVEEIVAPSVVSVAPSVEESV-----NVEE	557
Qy	466	DTPSNRIDLDFDSIAAQAQNGILSGALTQDEETQKRDADWNAIESTDSVYEPETFNPNP	525
Db	558	SVAEN-----VEESVAENVEESVAENVEESVAENVEEIVAPTVEEIVAP	601
Qy	526	VEIVIDTPEPESVAQTAENK-PETVDTDFSDNLPNNHIGTTEETASAKPASPGLAGFL	583
Db	602	TVEEIVAPSVVSVAPSVEESVEENVEESVAENVEESV-----	647
Qy	584	KASSPETILEKTVAEVOTPEELHDFLVKYETDVAETAETPTDFNAAADDLSALLQPAEA	643
Db	648	-AENVEEIVAPTVEEIVAP-TVEEIVAPSVVSVAPSVEESVEEN-VEESVAENVEESVA	704
Qy	644	PSVENITETVAETPDPFNATADDLSALLQPEVPAVEENAAEIVADDLSALLQPAEAPAV	703
Db	705	ENVEESVAENVEE-----SVAENVEEIVAPTVEEIVAPTVEEIVAPSVVSVAPSVEESV	759
Qy	704	EENVTETVAETSDPHTAADDLSALLQPAEVPVAENVTKTVAEIPDFNATADDLSALLQ	763
Db	760	EENVEESVAENVE-----ESVAENVEESVAENVEESVAPTVEE-----VAPSVEE	805
Qy	764	SEVPAVEENAAEITLETPTDSNTSEADALPDFLKQGEETVDMQSVLSSEINPNNADTSFP	823
Db	806	SVAPSVEESVA-----ENVATNLSQ-NLLSNLLGGIETEIEIKDSILNEIEEVKENVVTI-	859
Qy	824	SESVGSDAPSEAKYDLAEWYLEIGDRDAAEITVQKLEAEAGDVLKRA	871
Db	860	LENV-BETTAESVTTFSNILEIQENTITNDTIEEKLEHENVLSAA	906
XX	AAU96699;		
XX	07-AUG-2003 (revised)		
DT	30-JUL-2002 (first entry)		
XX	Plasmodium falciparum liver stage antigen-3 (LSA-3) #1.		
XX	Liver stage antigen-3; LSA-3; vaccine; Th1-inducing adjuvant;		
XX	malaria parasite; malaria; protein-specific cytotoxic T cell response;		
XX	gamma-interferon.		
XX	Plasmodium falciparum.		
XX	Key Location/Qualifiers		
XX	Region 1..278		

Db	311	VAENVVEESVAENVVEESVAENVVEESVAENVVEEIVAPTVEEIVAP-----TVEEIVAPSVV	364
QY	233	KA-NPNVSAHGRLRAGSVLHPIPNLNRKAEQKQPTAKPKAETASMPSEPSKQATVKPV	291
Db	365	ESVAPSVESVE-----ENVEESVAENVVEESVAENVVEESVAENVVEESV	415
QY	292	EKPEAKVAPE-----AKAEKPAVRPEVPAANTAETAESAQOAAASADITPTDET	346
Db	416	AENVEEIVAPTVEEIVAPTVEEIVAPSVVESVAPSVVEEENVEESVAENVVEESVAENV	475
QY	347	GNAYSEPVQVSAEETESGLFGGSYTLLLAGGAALIALLLLLRLAQSKRARTTESVP	406
Db	476	EESVAENVVEESVAENVVEES-----VAENVVEESVA	504
QY	407	EE-BPDLDADAADGIEITFAVETPATPEPAPKNDVNDTIALDGESEELSAKQTFDVEI	465
Db	505	ENVEESVAENVVEEIVAPTVEEIVAPTVEEIVAPSVVESVAPSVVEE-----NVEE	557
QY	466	DTPSNRIDLPDSLAAQNGILSALTDQDETOKRADADWNAIESTDSVVEPETFNPYP	525
Db	558	SVAEN-----VVEESVAENVVEESVAENVVEESVAENVVEEIVAPTVEEIVAP	601
QY	526	-VEITDTPPEPSVAQTAENK-PETVDTPDSNLPNNHIGTEETASAKPASPGLAGFL	583
Db	602	TVEEIVAPSVVESVAPSVVEESVAENVVEESVAENVVEESVAENVVEESV-----	647
QY	584	KASPETILKTVAVQVTPPELHDFLKVETDVAETAPETPDNFNAADDLGALLQPAEA	643
Db	648	-AENVEEIVAPTVEEIVAP-TVEEIVAPSVVESVAPSVVEEEN-VESVAENVVEESVA	704
QY	644	PSVEENITETVAETPDNFNATADDLSALLQSPVPAVEENAAEIVADLGSALLQPAEPAV	703
Db	705	ENVEESVAENVVEE-----SVAENVVEEIVAPTVEEIVAPSVVESVAPSVVEESV	759
QY	704	EENVETVAETSDPFHTAADLSDLSALLQPAEVPVAEENVTVKVAEIPDNATADDLSALLQ	763
Db	760	EENVEESVAENVE-----ESVAENVVEESVAENVVEESVAPTVEE-----VAPSVVE	805
QY	764	SEVPVAEENAAEITLTPDSNTSEADALPDFLKDGEETVDWMSIYLSEENIPNNADTSFP	823
Db	806	SVAPSVVESVA-----ENVATNLSL-NLLSNLLGGIETEIKDSILNEIEEVKENVVTTII-	859
QY	824	SESVGSDAPSEAKYDLAEWKVLEIGDRDAAAEVQVKLLREAGDVILKRA	871
Db	860	LENV-BETTAESTVTTFSENIIEIQENTITNDTIEKLEELHENVLSSAA	906
RESULT	15		
ABJ25647			
ID	ABJ25647	standard; protein; 1000 AA.	
AC	ABJ25647		
XX			
XX			
DT	16-APR-2003	(first entry)	
XX			
DE	Aspergillus fumigatus essential gene protein #305.		
XX			
KW	Fungicide; cytostatic; essential gene; Aspergillus fumigatus; infection;		
KW	cancer; contamination; biofilm; antibody; immune response.		
OS	Aspergillus fumigatus.		
XX			
FN	WC0200286090-A2.		
XX			
PD	31-OCT-2002.		
XX			
XX	23-APR-2002; 2002WO-US013142.		
PF			
PR	23-APR-2001; 2001US-0285697P.		
PR	27-APR-2001; 2001US-0287066P.		
PR	05-JUN-2001; 2001US-0295890P.		
PR	09-JUL-2001; 2001US-0203890P.		

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 2, 2004, 19:15:51 ; Search time 20 Seconds
(without alignments)
2271.543 Million cell updates/sec

Title: US-09-743-674-2

Perfect score: 4404

Sequence: 1 MPAGRLPRRCPPMWTFTDCT.....EBAEGDVLKRAQALAEGLG 880

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA.*

- 1: /cgn2.6/prodata/2/iaa/5A COMB.pep.*
- 2: /cgn2.6/prodata/2/iaa/5B COMB.pep.*
- 3: /cgn2.6/prodata/2/iaa/6A COMB.pep.*
- 4: /cgn2.6/prodata/2/iaa/6B COMB.pep.*
- 5: /cgn2.6/prodata/2/iaa/PCTUS COMB.pep.*
- 6: /cgn2.6/prodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	398.5	9.0	975	US-09-252-991A-28749	Sequence 28749, A
2	315.5	7.2	1786	US-08-973-462-8	Sequence 8, Appli
3	289	6.6	8991	US-08-714-741-32	Sequence 32, Appli
4	276.5	6.3	2468	US-09-976-594-726	Sequence 726, Appl
5	269	6.1	540	US-08-973-462-22	Sequence 22, Appli
6	267	6.1	688	US-09-141-047-8	Sequence 8, Appli
7	250.5	5.7	630	US-08-973-462-9	Sequence 9, Appli
8	236	5.4	703	US-09-252-991A-17865	Sequence 17865, A
9	236	5.4	1596	US-08-978-277A-4	Sequence 4, Appli
10	235	5.3	1346	US-08-635-121-2	Sequence 2, Appli
11	235	5.3	1346	US-08-978-277A-2	Sequence 2, Appli
12	235	5.3	2137	US-09-134-001C-4463	Sequence 4463, Ap
13	233.5	5.3	2482	US-09-252-991A-16967	Sequence 16967, A
14	232.5	5.3	1848	US-08-296-791-6	Sequence 6, Appli
15	232.5	5.3	1848	US-09-839-996-6	Sequence 6, Appli
16	232.5	5.3	1848	PCT-US95-10661A-6	Sequence 6, Appli
17	232.5	5.3	1848	US-10-080-505-6	Sequence 6, Appli
18	232.5	5.3	2315	US-09-543-681A-5434	Sequence 5434, Ap
19	230.5	5.2	1018	US-08-072-610-2	Sequence 2, Appli
20	230.5	5.2	1018	US-08-719-822B-2	Sequence 2, Appli
21	230.5	5.2	1018	US-09-092-458-2	Sequence 2, Appli
22	217	4.9	1231	US-08-904-263A-4	Sequence 4, Appli
23	217	4.9	1231	US-09-434-123A-4	Sequence 4, Appli
24	216.5	4.9	1075	US-09-252-991A-18387	Sequence 18387, A
25	214	4.9	928	US-09-134-000C-6590	Sequence 6590, Ap
26	213.5	4.8	1601	US-09-345-473E-40	Sequence 40, Appli
27	213	4.8	1702	US-08-296-791-5	Sequence 5, Appli

28	213	4.8	1702	4	US-09-839-996-5	Sequence 5, Appli
29	213	4.8	1702	4	US-10-080-505-5	Sequence 5, Appli
30	213	4.8	1702	5	PCT-US95-10661A-5	Sequence 5, Appli
31	212.5	4.8	1404	4	US-10-164-595-78	Sequence 78, Appli
32	212.5	4.8	2756	1	US-08-375-709-11	Sequence 11, Appli
33	212.5	4.8	2756	1	US-08-752-929-11	Sequence 11, Appli
34	212.5	4.8	2756	3	US-09-090-793-7	Sequence 7, Appli
35	212.5	4.8	2756	4	US-09-231-899-7	Sequence 7, Appli
36	211.5	4.8	3788	4	US-09-336-447A-76	Sequence 76, Appli
37	210.5	4.8	1140	4	US-07-757-022B-104	Sequence 104, App
38	210.5	4.8	1363	4	US-07-757-022B-52	Sequence 52, Appli
39	210.5	4.8	1404	4	US-07-757-022B-2	Sequence 2, Appli
40	210.5	4.8	1404	4	US-07-757-022B-62	Sequence 62, Appli
41	210	4.8	941	4	US-07-757-022B-14	Sequence 14, Appli
42	210	4.8	1022	4	US-07-757-022B-84	Sequence 84, Appli
43	210	4.8	1038	4	US-07-757-022B-74	Sequence 74, Appli
44	210	4.8	1049	4	US-07-757-022B-58	Sequence 58, Appli
45	210	4.8	1270	4	US-07-757-022B-44	Sequence 44, Appli

ALIGNMENTS

RESULT 1
US-09-252-991A-28749
; Sequence 28749, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 28749
; LENGTH: 975
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-28749

Query Match 9.0%; Score 398.5; DB 4; Length 975;
Best Local Similarity 23.1%; Pred. No. 1.8e-20;
Matches 240; Conservative 136; Mismatches 378; Indels 285; Gaps 44;

QY	14	TKFTDC---TRNRIOPTTHRGVILKNNRQIKLIASVAVASFOAHA-GLGLNIOQNL	69
DB	35	TASTALPLVKEKILKP--RDYTWRLTLVRAIAAASVLTSGMAHGLGLGELITLKSAL	92
QY	70	DEPFGSITVTGEEAKALGGGSVTVS-----BKGLTAKVHKL-----GDKA	111
DB	93	NQPLDAEIELL--EVRD-LGSEVPIPLASPEFSKAGVDRLYLTLDKFTPVKENGKS	149
QY	112	VIATVSEQAVRDPVLVFRIGA---GAQVREYTAIDLPVGYSPKTKSALSDGKTHKRTAP	167
DB	150	VIRVTSKPEQVQBPFLNVLQVLPNGRLIREYTLVLDPLPSYQAAASAPQAPV---SAP	206
QY	168	TAESQENQAKALRTDKDSANAIVK---PAYNGKTHVTKGETVKQIAAARPKHLTL	224
DB	207	RATG-----APRAPQAPVPTAPGSDTYRTV-SNDTLWEIAQRNRTRDRSV	254
QY	225	EQVADALLKANPNVSAHG---RLRAGSVLHINLRKAEQPKPOTAKPKAETAS-----	276
DB	255	PQAMLAFOELNFGAFVDGNINELKSGQVLRITPEQOMLERSREALSQVQAQNSWRGR	314
QY	277	-----MPSEPSK-QAT-----	286
DB	315	NPAAGSAGARQIDAQRNAAGSAPSKVDATDNLRLVSGEGKASKGADKGGKSKALADT	374

QY 287 -----VEKPVKPEAKVA----- 299
Db 375 LAVTKESLDSRRENEELQSRMQDLQSLQDKLQKLIQKDAQAKLQGLQAGQGAQAP 434
QY 300 -----APAKAEKPAVRPEVPVPAANTAASETAESAPOEAAASALDITPDE--TCNA 349
Db 435 NAALFDASQPNAAQAQAPQGTAAAAAPTTPAGEAPAAPAAQPPVAPPAAEKPPAPA 494
QY 350 VSEPEVQVSAEETESGLFGGSYTLILAG-----GGAALIALLLRLAQSRRARTE 402
Db 495 VPAPVQAAEQAPSEI-----DELLANPLWLAVIGGSALLLVLLMILSRRAQKEK 549
QY 403 ESEV-----EEBPDLDAAADGI-EITFAEVETPATPEPAPKNDVNDTLALDGESEEE 454
Db 550 EEVQAFADAGEEQEDALDKGQFDDITLDEPE-PQVAAAAPOVEKTTAQTSDALGEAD 608
QY 455 L-----SAKOTFVETTPNSNRIDLDFDSLAA-AQNGILSGALTODEETOKRADAD 504
Db 609 IYIAYGRFNOAAELLQNAIYDEP-QRTDLRLKLMVVAEMGDREGFARQENELREIGGAQ 667
QY 505 WNAIESTDSVYEPETFPNPNVEIIVDTPEPESVAQTAENKPEVDTDFSNLPSNNHIG 564
Db 668 -PQVEQLKSY-----PAMVAVAAVAGLAGKLAQDELSFSLD-DLSLDSGH-- 714
QY 565 TEETASAKP-ASPSGLAGLAKASSPETTILEKTVAEVQTPPE-ELHDFLKVYETDAVAETAP 622
Db 715 -----AAKPDAAQGLDDAFDLSLDLDGDDVQADLKSDSGALDDLTLDSLDLAASAA 769
QY 623 ETPFNAAAADLSALLQPAAPSVEENTETVETPDENATADLSALLQSEVPAPVEN 682
Db 770 DKP-----VDBLDFGLDFAE-----LAETPS-QPKHDDLGDFSLLDADP--BDK 810
QY 683 AAEIVADLSALLQ-----PAEAPAVEENVETVETVETSDPHTAAD--DLSALLQPAEVPAVE 737
Db 811 LSD-----DFLLSLNDEVPAAPADNEFTLDEAAEPAISLPDDFDLSLADEPTEPAPE 867
QY 738 ENVTKTVAEIPDFNATADLSALLQSEVPAPVEENAAITLETPTSNT---SEADALPDF 794
Db 868 KGEDSFAQLDEVSQAQLDELAS-----NLDEPKSATPSFAEDAAVAS 910
QY 795 LKDGEEVVDWSIYLSEENIPNNAADTSPSEVSGDAPSEAKYDLAEMYLEIGDRDAAB 854
Db 911 ALDGD-----ADDDFDLFGADEAAT--KLIDARAYIDMGDSEGARD 950
QY 855 TVQKLEEAEGDVLKRAQA 873
Db 951 ILDEVL--AEGNDSQAAEA 967

RESULT 2
US-08-973-462-8
; Sequence 8, Application US/08973462B
; Patent No. 6191270
; GENERAL INFORMATION:
; APPLICANT: DRUILHE, PIERRE
; APPLICANT: DAUBERSES, PIERRE
; TITLE OF INVENTION: MALARIAL PRE-ERYTHROCYTIC STAGE POLYPEPTIDE MOLECULES
; FILE REFERENCE: 0660-0125-0 PCT
; CURRENT APPLICATION NUMBER: US/08/973,462B
; CURRENT FILING DATE: 1998-02-06
; EARLIER APPLICATION NUMBER: PCT/FR96/00894
; EARLIER FILING DATE: 1996-06-12
; EARLIER APPLICATION NUMBER: FR 95/07007
; EARLIER FILING DATE: 1995-06-13
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 8
; LENGTH: 1786
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Polypeptide
US-08-973-462-8

Query Match 7.2%; Score 315.5; DB 3; Length 1786;
Best Local Similarity 20.2%; Pred. No. 5e-14;
Matches 167; Conservative 147; Mismatches 331; Indels 183; Gaps 26;
QY 64 NIQNLDEPFGSGITVTGEEAKALLGGSVTVSEKGLTAKVHKLGDKAV---IAVSSQA 120
Db 242 NVEENVEENDGVSASSVEES-----IASSVDESIDSSEENVEENVAPTVEEI 286
QY 121 VRDVLVFRICAGAVREYTAILDPGVGSPKTSALSADGKTHRTKTAETAS--QENQNAK 178
Db 287 VAPSV-----ESVAPSVSEESVEENVEES 310
QY 179 ALRXTDKKDSANAAVKPAYNGKTHVRKGE-----TVQIAAAIRPKHLTLEQ--VADALL 232
Db 311 VAENVEESVAENVEESVAENVEEIVAPTVEEIVAP-----TVEEIVAPSV 364
QY 233 KA-NPNVSAHGRRLRAGSVLHLPNLNRKAEQPKQTKAPKAEATASMPSEPSKQATVRKPV 291
Db 365 ESVAPSVSEESVE-----ENVEESVAENVEESVAENVEESVAENVEESV 415
QY 292 EKPEAKVAAPR-----AKAEKPAVRPEVPVPAANTAASETAESAPOEAAASALDITPDET 346
Db 416 AENVEEIVAPTVEEIVAPTVEEIVAPSVSVESVAPSVSEESVAENVEESVAENV 475
QY 347 GNAYSEPVQVSAEETESGLFGGSYTLILAGGGAALIALLLRLAQSRRARTEESP 406
Db 476 EESVAENVEESVAENVEES-----VAENVEESVA 504
QY 407 EE-EPDLDAAADGGLITFAEVETPATPEPAPKNDVNDTLALDGESEBELSAKOTFVET 465
Db 505 ENVEESVAENVEEIVAPTVEEIVAPTVEEIVAPSVSVESVAPSVSEESVEE-----NVEE 557
QY 466 DTPSNRIDLPDLSAAAQNGILSGALTODEETOKRADADWNAIESTDSVYEPETFPNPN 525
Db 558 SVAEN-----VEESVAENVEESVAENVEESVAENVEEIVAPTVEEIVAP 601
QY 526 -VEIVIDTPESPESVAQTAENK- PETVDTDFSDNLPSNNHIGTEETASAKPSPGLAGFL 583
Db 602 TVEEIVAPSVSVESVAPSVSEESVAENVEESVAENVEESVAENVEESV----- 647
QY 584 KASSPETTILEKTVAEVQTPPEELHDFLKVYETDAVAETAPETPDFNAAAADLSALLQPAEA 643
Db 648 -AENVEEIVAPTVEEIVAP-TVEEIVAPSVSVESVAPSVSEESVAENVEESVA 704
QY 644 PSVEENITETVAETPDFNATADLSALLQSEVPAPVEENAAITVADLSALLQPAAPAV 703
Db 705 ENVEESVAENVEE-----SVAENVEEIVAPTVEEIVAPTVEEIVAPSVSVESVAPSV 759
QY 704 EENVETVAETSDPHTAADDLSALLQPAEVPAPVEENVTKTVAEIPDFNATADLSALLQ 763
Db 760 EENVEESVAENVE-----ESVAENVEESVAENVEESVAPTVEEI-----VAPSV 805
QY 764 SEVPAPVEENAAITLETPTSNTSEADALPDFLKDGEETVDWSIYLSEENIPNNAADTSP 823
Db 806 SVAPSVSEESVA-----ENVATLSD--NLLSNLGGTETETIKDSILNEIEEVKENVVTIT- 859
QY 824 SESVGSADSPSEAKYDLAEMYLEIGDRDAABAAATVQKLEEAEGDVLKRA 871
Db 860 LENV-EETTAESVTTFSNILEEIQENTITNDTIEBKLEELHENVLSAA 906
RESULT 3
US-08-714-741-32
; Sequence 32, Application US/08714741
; Patent No. 6500613
; GENERAL INFORMATION:
; APPLICANT: Briles, David E.
; APPLICANT: McDaniel, Larry S.
; APPLICANT: Swiatlo, Edwin
; APPLICANT: Yother, Janet
; APPLICANT: Crain, Marilyn J.
; APPLICANT: Hollingshead, Susan

APPLICANT: Tart, Rebecca
 APPLICANT: Brooks-Walter, Alexis
 TITLE OF INVENTION: PNEUMOCOCCAL GENES, PORTIONS THEREOF,
 TITLE OF INVENTION: EXPRESSION PRODUCTS THEREFROM, AND USES OF SUCH GENES,
 TITLE OF INVENTION: PORTIONS AND PRODUCTS
 NUMBER OF SEQUENCES: 47
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Curtis, Morris & Safford, P.C.
 STREET: 530 Fifth Avenue
 CITY: New York
 STATE: New York
 COUNTRY: U.S.
 ZIP: 10036
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/714,741
 FILING DATE: 16-SEP-1996
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: Frommer Esq., William S.
 REGISTRATION NUMBER: 25,506
 REFERENCE/DOCKET NUMBER: 454312-2460
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (212) 840-3333
 TELEFAX: (212) 840-0712
 INFORMATION FOR SEQ ID NO: 32:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 8991 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: amino acid
 US-08-714-741-32

Query Match 6.6%; Score 289; DB 4; Length 8991;
 Best Local Similarity 21.6%; Pred. No. 4.8e-11;
 Matches 198; Conservative 92; Mismatches 313; Indels 312; Gaps 34;
 QY 146 VGYSPKTKALSQKTHRTAFTAESQENONAKALRT--DKKDSANAAVKPAYNGKTHT 203
 DB 7741 VEYFKT-----DAEQTEQYLAARXDLADKKALEKTEADLKAVNEPEKPAETPAPA 7795
 QY 204 VRKGETVKQIAAIRPK-----HLTLEQVADALLKANPNVSAHGRLAGSVLHLPNIN 256
 DB 7796 PKPEQPAEQPKPAPAPQAPAPAPKPEKTDQQA-----EDYARRSEEEYNRLPQQ 7846
 QY 257 RIKAEQPKP-----QTAQPKAETASMPSEPSKOATVEKPKPEAKVAAP 301
 DB 7847 PKPAEKPAAPKPEQVPVAPXENPAPAPKAPAPQPLKPEEPAEQPKP-EKPEEPAGQP 7905
 QY 302 E-----AKAEKPAVRP-----EPVPAANT-----320
 DB 7906 EPEKDDQAGEDIYARRSGEYNNRFPQOPPKAEKPAAPKPEQVPVAPKTLKAKLAG 7965
 QY 321 AASETAESA-----PQEAASADTPTDETGNV 350
 DB 7966 AKSKATKAAELEPELEKAEAELENLLSTDPGKTQDELDEKAEAELENKVEALPNQV 8025
 QY 351 SEPVEQVSAEETESLFGSGVTLTLAGGGAALIALLLLRLLAQSKRARRTEESVPEEP 410
 DB 8026 SELEELSLELNLKDAETNNVEDYIKG-----LEAIAIKQAELEKTPKELDA 8075
 QY 411 DDDAAADDGIEITFAEVETPA-----TPEAPKNDVN-DTIALDG-----ESEER- 454
 DB 8076 ALNELGPDGDEBETPPPEAPAEQPKPEKPAETPAPAPKPEKSADQAEEDYARRSEERY 8135
 QY 455 --LSAKQTFDVTDTPS-----NRIDDFDSL-----479
 DB 8136 NRLTQOPPKAEKPAAPAPKPEQAPAPAPKPSRGLATKKLNLAERAIELLKLLGLEPGL 8195

QY 480 --AAQNGILSGAL-----TQDETQKRADWNNAIESTDSVVEPETFNPVPIVDT 532
 DB 8196 EKAGAGLGNLSTLDPGKTDDELDEKAEAELEN-----KKVEAL-----8235
 QY 533 PEPESVAQTAENKPTVDTFSDNLPNNHIGTEETASAKPASPGLAGFLKASSPETIL 592
 DB 8236 --PNQVALEEL-----SKLEDNL-----KDAETNHVEDYIKGLEEAIA 8274
 QY 593 EKTVAEVQTPTEELHDFLKVYETDAVETAPETDFNAAADDLSALLQAPAPAPSVEENITE 652
 DB 8275 TKQAELEKTPKELDAALNELGPDGDEE---ETPAEA-----PAQPKPEKPAEE 8321
 QY 653 TVAETPDFNATADDLSALLQSPVPAVVENAAEIVADDLSALLQ-----PAEAPAVE 704
 DB 8322 TPAPAPKPEKSADQ-----QAEDYARRSEEEYNRLTQOPPKAEKPAAPAPK 8370
 QY 705 ENVTVAETSDFTTAADDLSALLQPAEVPVAVENVT-----KTVAEIPDFNA 752
 DB 8371 PE-----QAPAPKPKQKVNLENLLSTLDPGKTDDEL-DKGA 8407
 QY 753 TADDLSALLQ--PSEVPVAVENAAEITLETPTSNTSEADA--LPDFLKQGEETVDSIY 808
 DB 8408 AEALNKKVEALEPNVXELSEE-----LSPEDNLKDAETNHVEDYIKGLEEAIAIKQA 8462
 QY 809 LSEEN-----IPNNADTSFPPSESVGSDAPSEAKYDLAEMYLEIGDRDAAAEIVQ 857
 DB 8463 ELEETPOEVDALNDLVPDGGEBETPAPAPQDPAPAPAPNAEQPAPAPKPEKSAD---8519
 QY 858 KLEEAEGDVLKRAQ 872
 DB 8520 ---QAEEDYARRSE 8531

RESULT 4

US-09-976-594-726
 ; Sequence 726, Application US/09976594
 ; Patent No. 6673549
 ; GENERAL INFORMATION:
 ; APPLICANT: Furness, Michael
 ; APPLICANT: Buchbinder, Jenny
 ; TITLE OF INVENTION: GENES EXPRESSED IN C3A LIVER CELL CULTURES TREATED WITH STEROIDS
 ; FILE REFERENCE: PA-0041 US
 ; CURRENT APPLICATION NUMBER: US/09/976,594
 ; PRIOR FILING DATE: 2001-10-12
 ; PRIOR APPLICATION NUMBER: 60/240,409
 ; NUMBER OF SEQ ID NOS: 1143
 ; SOFTWARE: PERL Program
 ; SEQ ID NO 726
 ; LENGTH: 2468
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: misc feature
 ; OTHER INFORMATION: Incyte ID No. 6673549 4295277CD1
 US-09-976-594-726

Query Match 6.3%; Score 276.5; DB 4; Length 2468;
 Best Local Similarity 20.1%; Pred. No. 5.6e-11;
 Matches 185; Conservative 125; Mismatches 333; Indels 277; Gaps 35;
 QY 150 PKTKALSQKGT--HRKTAPTAESEONAKALRTDKDSANAAVKPAYNGKTHTVRK 207
 DB 656 PKPEVAKEDKTPFKKEPKKEEYKVEKKEIKKEEKEPKKEKVEKKEKVEK 715
 QY 208 ETVKQIAAIR--PKHLTLEQVADALLKANPNVSAHGRLAGSVLHLPNINRLKAEQPKQ 266
 DB 716 EEKVKVEEKEPKKEIKKLKPKAKSSTP-----LSBAKDA 753
 QY 267 TAKPKAETASMPSEPSKOATVE-----KPKPEKPAKVAAPAEKAEKPAVPEPVAAN 319
 DB 754 ALKPKV-----PKKEESVKKDSVAAGKPKKEKIKVKEKGAEE-AVAAAVGTGAT 804

```

QY 320 TAASETAESA-----PQAAAAAIDTPTDTCNAVSEPVQVSAEEETSGLFGGYSY 373
D 805 TAAMAAAGIAIAPAKELERSLMSPEDLTDFELKAEVDVTKDKP----- 856
QY 374 LLLAGGGAALIALALLLAQSKARTEESVPEEPDLD-----DAADDGIEITFAEVET 429
D 857 -----QLELIBDEEKLKETEPEVAIVIKEREVTKGPAESPDEGITTTEGEGEC 905
QY 430 PATP-----EPAPKNDVNTLAL-----DGE----- 450
D 906 EQTPEELEPVKQGVDDIEKPEDEGAGFESESTGDYEEKAEETAEAPEDGSEHVCS 965
QY 451 -----SEELSAKQTFDV-----ETDTSNRIDLDPSLAAQNGILSCALTQDBET 497
D 966 ASKHSPTDEESAKAEDAYIREKRESVASGDRAEDMDZ--ALIEG-----BAQS :016
QY 498 QKRADADNNAIESTDSVYPTFPNPNPVEIVD-----TPESPVAQT 541
D 1017 EEEADEEDKAEADREVEPEKMAEDYVMAVDKAAEAGAEQYGLTTPTKQLGAQS 1076
QY 542 AENKPEITDTPSDNLPNNH-----IGTEETASAKPASPSGLAGFLKA----- 585
D 1077 PGRFPASSIHD--ETLPGGSESEATASDEENREDQPREFTATSGTQSTIIEISSEPTMD 1134
QY 586 --SSPETILEKTVAEVQTPPELHDFLKV--YETDAVAE---TAPETP----- 625
D 1135 EMSTPRDVMSETNNEETESQEFVNITKYESLSYQSEYKPADVTPLNGFSEGSKTD 1194
QY 626 ----DFNAAA-----DLSAL-----LOPA 641
D 1195 TDGKDYNASASTISPPSMEEEDKFSRALSADYCVSEVKASTTLDIKDISAVSSKVSFS :254
QY 642 EAPSEVENITETVAETP-----DFNATADDLSALLQSEVPAAVEENAAIVADD-LSAL 694
D 1255 KSPSLSPSPSPPLEKTPLGERSVNFSLTPNEIKVSAEAEVAPVSEVQVVEEHCAEP 1314
QY 695 LQAPAEAPAEVENVETVAETSDFTAAADLSALLOPAEV-----PAV-----ENNV 740
D 1315 DKLEWVSPQSQTGAGHTPYQSPTEKSHL--PTEVIEKPPAVPVSPFSDAKDENE 1373
QY 741 TKTV-----EIPDFNATADDLSALLQSEVPVAV--EENAAEITLTPTDNTSEADALPDF 794
D 1374 RASVSPMBEPVDP-----SESPIEKVLSPRLSPPLIGSESAYESFLSADDKASGRGAESPFE 1430
QY 795 LKQGEETVDKGIYILSEENIPNNAADTSFSPESVSGSDAPSEAKYDLAEMYLEIGDR----D 850
D 1431 EKSGKGSPDQVSPVSE-----MTSTSLYQDKQEGKSTDFAPIKEDFGQEKKTDD :480
QY 851 AAAETVQKLL-----EEAEGDV 867
D 1481 VEAMSSQPALADDERKIGDV 1500

```

RESULT 5

```

US-08-973-462-22
; Sequence 22, Application US/08973462B
; Patent No. 6191270
; GENERAL INFORMATION:
; APPLICANT: DRUIIHE, PIERRE
; APPLICANT: DAUBERSTES, PIERRE
; TITLE OF INVENTION: MALARIAL PRE-ERYTHROCYTIC STAGE POLYPEPTIDE MOLECULES
; FILE REFERENCE: 0660-0125-0 PCT
; CURRENT APPLICATION NUMBER: US/08/973,462B
; CURRENT FILING DATE: 1998-02-06
; EARLIER APPLICATION NUMBER: PCT/FR96/00894
; EARLIER FILING DATE: 1996-06-12
; EARLIER APPLICATION NUMBER: FR 95/07007
; EARLIER FILING DATE: 1995-06-13
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 22
; LENGTH: 540

```

```

; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Polypeptide
US-08-973-462-22

Query Match 6.1%; Score 269; DB 3; Length 540;
Best Local Similarity 20.9%; Pred. No. 2e-11;
Matches 131; Conservative 112; Mismatches 262; Indels 122; Gaps 19;

QY 166 APTAES--QENQNAKALRKTDKDSANAAVKPAYNGTKHTVRKGE-----TVKOIAAAIR 218
D 18 APSVEEVEEVEEVEEVEEVEEVEEVEEVEEVEEVEEVEEVEEVEEVEEVEEVEEVEEVEE 75
QY 219 PKHITLQ--VADALLKA--NPNVSAHGRLAGSVLHIPLNLRKIAKQPKQPTAKPKAETAS 276
D 76 ----TVEIVAPSVVSVAPSVSEVE-----ENVEEVAEVEEVEEVEEVEEVEEVEEVEEVA 122
QY 277 MPSEPSQATVKEKPEKAAPE-----AKAEKPAVRPEPVPAANTAASETAESAP 331
D 123 ENVEEVAEVEEVEEVEEVEEVEEVEEVEEVEEVEEVEEVEEVEEVEEVEEVEEVEEVEEVE 182
QY 332 QERASAIPTDPTDTCNAVSEPVQVSAEEETSGLFGGYSYTLILLAGGGAALIALLLILR 391
D 183 ESVAEVEEVEEVEEVEEVEEVEEVEEVEEVEEVEEVEEVEEVEEVEEVEEVEEVEEVEE 216
QY 392 LAOSKRARTEESVPEE--EPDLDDAADGIBITFAEVETPATPEKPNKNDVNTLALDGE 450
D 217 ----VAEVEEVEEVEEVEEVEEVEEVEEVEEVEEVEEVEEVEEVEEVEEVEEVEEVEEVEE 271
QY 451 SEELSAKQTFDVTPTSDNLPNNH-----DLSAL-----LOPA 641
D 272 SVEE-----NVESVAE-----VEEVAEVEEVEEVEEVEEVEEVEEVEEVEEVEEVEE 308
QY 511 TDSVVEPETFPNPNP-VEIVDTPESVAQTAENK-PETVDTDFSDNLPNNHIGTBEET 568
D 309 VEEIVAPTVEEIVAPTVEEIVAPSVVSVAPSVSEVEEVEEVEEVEEVEEVEEVEEVEEVEE 368
QY 569 ASAKPASPSGLAGFLKASPTILEKTVAEVQTPPELHDFLKVYETDAVAETAPTPDFN 628
D 369 V-----AENVEEIVAPTVEEIVAP--TVEEIVAPSVVSVAPSVSEVEEVEEVEEVEEVEE 412
QY 629 AAADLSALLQPAEAPSEVENITETVAETPDNATADDLSALLQSEVPAAVEENAAEIVA 688
D 413 --VEEVAEVEEVEEVEEVEEVEEVEEVEEVEEVEEVEEVEEVEEVEEVEEVEEVEEVEEVE 466
QY 689 DDLALLQPAEAPAEVENVETVAETSDFTAAADLSALLQPAEVPAAVEENVTKTVAEIP 748
D 467 PSVVEEVAEVEEVEEVEEVEEVEEVEEVEEVEEVEEVEEVEEVEEVEEVEEVEEVEEVEE 520
QY 749 DFNATADDLSALLQSEVPAAVEENAAE 775
D 521 -----VAPSVVEEVAEVEEVAE 539

```

RESULT 6

```

US-09-141-047-8
; Sequence 8, Application US/09141047A
; Patent No. 6043085
; GENERAL INFORMATION:
; APPLICANT: Walker, David H.
; APPLICANT: Yu, Xue-Jie
; TITLE OF INVENTION: Ehrlichia canis 120-kDa Immunodominant Antigenic
; FILE REFERENCE: D6143
; CURRENT APPLICATION NUMBER: US/09/141,047A
; CURRENT FILING DATE: 1998-08-27
; NUMBER OF SEQ ID NOS: 15
; SEQ ID NO 8
; LENGTH: 688
; TYPE: PRT
; ORGANISM: Ehrlichia canis
; FEATURE:

```


Best Local Similarity 22.6%; Pred. No. 7.7e-09;
Matches 184; Conservative 104; Mismatches 273; Indels 254; Gaps 42;

```

QY 40 QIKLIAAASVAAASQAHAGGNIQSNLDEPPSGSITVT-----CEBAKALGGGSV 93
Db 29 RVMLCAMSCLLAGQASA-LGVGDITLHSAHQDADIELLDVGLGDADEIEVRLAGADV 87
QY 94 TVSBKGLTAKVHKL-----GDKAVIASSEBQAVRDPVLFVRIGA-----G 133
Db 88 FAA-----AGVERLOFLNELRFSPVLQGRGGR--IHVSSIRPVQEPVLFIVEVAPNG 140
QY 134 AQVREYTAILDPVGVSPKTSALSDCKTKRKTAPTAEQENQNAKALRKTUKKUSANAAY 193
Db 141 RIVREFVLLDPLGTYPRMLPAARSGIEPQR-----QSSTPAPAPR-----SAAAVV 187
QY 194 KPAY--NGKTHTVRKGETVKQIAAIR-PKHLTLEOVADALLKANP-----NVSAGRUR 246
Db 188 DPALLEPDEYLARNQNLMAISGLRGAGNADRAQLMEALYQLNPQAFVADRH-LRKA 246
QY 247 GSVLHPIPLNRKAKQPKPTAK-----PKAETASMPSEPS-----KQATVE----- 288
Db 247 GVRLRLP--AGQPERGAPGAVKEAAVEVLPADAADVENAPAAALVEAQROADAEEALA 304
QY 289 -----KPVKPKAKVAAPEAK-AEKPAVRP-EPVPAANTAASET 325
Db 305 KQRELSQRMDDLQRLQALQQLQORDHQAELQOQLARRQAVRPAAPPPAA----- 357
QY 326 AAESAQAASAIPTDPTETGNVSEPVQVSEETESGLFGGSYTLILAGGGAALIA 385
Db 358 ---AAP--SVAQPVETPTD-----SQWRMIVLLV--LALLG 389
QY 386 LLLLRLAQSKRARTEESVPEEPDLDADGDIITFAEVETPATPEPAPKNDVNDTL 445
Db 390 VLLL-----RRREEAPVQAVEPKRVA---LNLPLRRAPRPPAAAPAK----- 432
QY 446 ALDGESEELSNAQTFDVTETPSNRIDLDFSLAAQNGILSGALTODEETQKRDADW 505
Db 433 -VEEQARPVAA-----PSSPPPPPPAPAAAPRAAMAAA----- 466
QY 506 NAIBETSDSYEPETENPNPV-----EVIDPPEPSVAQTAENKPEVTDTDFSDNLP 560
Db 467 DKLDGAD-IY--IAYRGVQARDLLRQVLAEQPRLS---ARMKLLVLVAELGD----- 514
QY 561 NHIGTETASAKPSPGLAGLAKASSPETILEKTVAEVQTPPEELHDFLKVYETDAVET 620
Db 515 -----AAGE-----DALAETLAGGNGSEAIDELGRYPV-LIQMP 549
QY 621 ABETDFNAADLALLOPAEAPSVEENITETVAETDFNATAD-----DLSALL 671
Db 550 ATETPAATTKDDWSD-LPLAEP-----VLQQPDATSGADGFGDLNLDLDWGALE 599
QY 672 QPSEVPVAEENAAEIVADLALLOPAEAP--AVEENYTE--TVAEISDEHTAADDLSAL 727
Db 600 NPLDNPDLPRRA-----ACKAEPAEELPAPESNLHELDPVAEYR--HLELQO----- 646
QY 728 LQPAEVPVAEENVTKVTAIPDNFATADLQ 762
Db 647 PEPATVPPEASASLDRACIDSGDLQASRIIL 681

```

RESULT 9

US-08-978-277A-4
Sequence 4, Application US/08978277A
Patent No. 6582956
GENERAL INFORMATION:
APPLICANT: Gelman, Irwin H.
TITLE OF INVENTION: TUMOR SUPPRESSOR GENE
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:
ADDRESSEE: Brumbaugh, Graves, Donohue & Raymond
STREET: 30 Rockefeller Plaza
CITY: New York
STATE: NY

```

; COUNTRY: USA
; ZIP: 10112-0228
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/978,277A
; FILING DATE:
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/665,401
; FILING DATE: 18-JUN-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Clark, Richard S
; REGISTRATION NUMBER: 26,154
; REFERENCE/DOCKET NUMBER: A30558 - 165/34008
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-408-2558
; TELEFAX: 212-765-2519
; TELEX:
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1596 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: internal
; ORIGINAL SOURCE:
; US-08-978-277A-4

```

Query Match 5.4%; Score 236; DB 4; Length 1596;
Best Local Similarity 20.2%; Pred. No. 2.6e-08;
Matches 168; Conservative 106; Mismatches 281; Indels 277; Gaps 37;

```

QY 258 IKAEQPKQTAQKPAETASMPSEPSKQATVTK-PVEKPE----- 295
Db 537 IHTSPESADQKQESSASSPEETTCLEKGPLEAPQDGEABEGTSDGKKREGITP 596
QY 296 ---AKVAAPAKAKPAVRPEPVPAANTAASETAAESAPOAASAIPTDPTETGN 351
Db 597 WASFKMVTPKRVRP-----SESDKEE-ELEKVKSATLSLSDTVSEMQ 641
QY 352 EPVEQVSAEETESGL-----FGGSYTLILAGGGAALIALLLLRLAQSKRARTEESVPE 407
Db 642 DEVKTGVGEQPEEPKRRVDTSVSWEALICVG-----SSKKRAKASSDDE 688
QY 408 -----BEPOLD-DAADGGIETFAE---VETPATPEPAPKNDVNDTLALDGE 450
Db 689 GGPRTLGGDSHRAEASKDKAAGTDAVPASTQEQDQAQSSSPFA-----GSP 737
QY 451 SE-BELSAKQTFD-VETDTPSNRIDLDFSLAAQNGILSGALTODEETQKRDADW 508
Db 738 SEGEVSTWESFKRLVTPRKKSXKLE---EKAEDSSVEQLSTEIPEPSREES---WVSI 790
QY 509 ES-----TDSVYEPETFNPNPVEIIVDTPEPESVAQTAEE---NKPET----- 548
Db 791 KKFIPGRRKKRAGKQEQATVEDSGFVEINDDPNVPVAVPLSEVNAVEREKWEAQGNT 850
QY 549 -----VDTDFSDNLPNNHI-----GTEETASAKPASPSGLAGFL-----KAS 586
Db 851 LPQLLGAVVSEELSKTLVHTVSVAVIDGTRAVTSVEERSPSWISASVTEPLEHTAGAM 910
QY 587 SP-ETILEKTVAEVQTP-----EELHDFLKVYETDAVAA--TAPETD----- 626
Db 911 PPVEEVTEKDIITAEETPVLTQTLPGKQADHDMVTSEVDFVSEAVTATETSEALTEV 970
QY 627 ----FNAAADDLSALLOPAEAPSVENITET-----VAETPD-----FNATADDLS 668

```


Db 971 EASGARETTDMVSAVSQLTSDPTTEBATPVQEVESGVLDTEEBERQTQALQAVADKVK 1030
QY 669 ALLQPSVPA-----VEENA----- 683
Db 1031 ---ESQVPAQTQVGTGSKALEKVEVEEDSEVLASEKQVMPKGPVQVAGAHHAQ 1087
QY 684 -----AEIVAD-----DLSALLOPAPAEVAVENVTETVAETSDPHTA 720
Db 1088 SETGQATPESLEVPEVADVHVATQVVKLQQLMEQAVAPESSETLTDSETNGSTPLAD 1147
QY 721 ADDLSALLOPAPAEVAVENVTETVAE---IPDNATADDLSALLOPSEVAVENAAEIT 777
Db 1148 SDTADGTQDETDSQSKATAAARQSVQTEEAATAKEEPSTLPNNVPAQEEHGE 1207
QY 778 LET--PDSNTSEADALPDFLKD--GEETVDWMSIYLSEENI-----PNN---AD 819
Db 1208 RDVLEPTQQLTAAPVPLAKTEVGQGEVDW---LDGEKVKBEQVVFVHSGPNSQKAAD 1264
QY 820 TSPFSESVG--SDAPSEAKYDLAEWYLEIGDRDAAAE-----TVQKLEAE 864
Db 1265 VTVDSEVMGVAGCQEKESQVSLSEEGEMETDVEKEKRETKPEQVSEBGE 1316

RESULT 10

US-08-635-121-2

; Sequence 2, Application US/08635121

; Patent No. 5910442

; GENERAL INFORMATION:

; APPLICANT: Gelman, Irwin H.

; TITLE OF INVENTION: TUMOR SUPPRESSOR GENE

; NUMBER OF SEQUENCES: 2

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Brumbaugh, Graves, Donohue & Raymond

; STREET: 30 Rockefeller Plaza

; CITY: New York

; STATE: NY

; COUNTRY: USA

; ZIP: 10112-0228

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: DOS

; SOFTWARE: FastSeq Version 1.5

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/635,121

; FILING DATE: 19-APRIL-1996

; CLASSIFICATION: 514

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER:

; FILING DATE:

; ATTORNEY/AGENT INFORMATION:

; NAME: Clark, Richard S

; REGISTRATION NUMBER: 26,154

; REFERENCE/DOCKET NUMBER: A30558 - 165/33603

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 212-408-2558

; TELEFAX: 212-765-2519

; TELEX:

; INFORMATION FOR SEQ ID NO: 2:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1346 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: peptide

; HYPOTHEICAL: NO

; ANTI-SENSE: NO

; FRAGMENT TYPE: internal

; ORIGINAL SOURCE:

US-08-635-121-2

Query Match

Best Local Similarity 5.3%; Score 235; DB 2; Length 1346;

Pred. No. 2.4e-08;

Matches 171; Conservative 102; Mismatches 268; Indels 304; Gaps 37;
QY 258 IKAEOQPKQTAQKPAETASMPSPSKOATVEK-PVEKPEAKVAPEAKPAVPPPP-- 314
Db 151 IHTESPESADEQKESASSPPEPTTCKLEKPLEAPRM-----GKLKELLRGEKKR 204
QY 315 -----VPAANTAASETA---AESAPO---EAAASADTPTDETGNVSPVEQVSAEE 362
Db 205 KDHSGLGLQKDGDTQETVRPSPESDKEELEKVKSATLSSTSTVSMQDEKVTGGEQK 264
QY 363 TE-----SGLFGSYLLLAGGGAALIALLLLRLA 393
Db 265 PEPFKRRVDTSVSWEALICVSSKKRARKASSDIRGPTL---GGGQ-----S 310
QY 394 QSK-----RARTTESVPPEEPDLDDAADGCIETFAVEVTPATPEPA----- 436
Db 311 QSRGQGRQORSRTDAVPASTQEQDQA-----QGSSSPPEPAGSPSEGEVSTWE 358
QY 437 -----PKNDVNDTLALDGESEELSAKOTFVETDTP--SNRID--LDFDSLAAQNG 485
Db 359 SFKRLVTPRKKSCKL-----EKEAGRTLVGACPLRSNRVKNLGFPLRNSPDG 411
QY 486 ILSGALTQDEETQGRADADWNAIESTDSVYEPETFPNPNFVPEIVDTPEPESVQAEN- 544
Db 412 GRKGQMGROBQA-----TVEDSGPVEINDEPDPVAVVPLSEYD 450
QY 545 --KPEIVDTDFSDNLPD-----NNHI-----GTEETASAKPAPSGLAGF 582
Db 451 AVEREKMEAGQNAELFCWGVSVSELSKTLVHTVSVAVIDGTRAVTSVEERSPWSAS 510
QY 583 L-----KASSP--ETILEKTVAEVQTP-----EELHDFLKVYETDAVE--TA 621
Db 511 VTEPLSHTAGAMPVVEVTEKDIIEETPVLTQTLPEGKADHDDMTSEVDTSEAVTA 570
QY 622 PETPD-----FNAADDLSALLQAPAPSVVEENITET-----VAETPD----- 659
Db 571 TETSEALRTEVTEASGABETTDWSAVSQLTSDPTTEATPVQVEGGVLDTEEBERQ 630
QY 660 -----FNATADDLSALLQPSVPA-----VEENA----- 683
Db 631 TQALQAVADKVK---EESQVPAQTQVORTGSKALEKVEVEEDSEVLASEKEKDVMPKG 687
QY 684 -----AEIVAD-----DLSALLOPAPAPAVEENV 707
Db 688 PVQEGAGHLAGQSGTQATPESLEVPETADVHVATQVVKLQQLMEQAVAPESSETL 747
QY 708 TETVAETSDPHTAADLSALLQAPAEVAVENVTETVAE---IPDNATADDLSALLQPS 764
Db 748 TDSNGSTPLADSDTADGTQDETDSQSKATAAARQSVQTEEAATAKEEPSTLPN 807
QY 765 EYPAVEENAAETLET--PDSNTSEADALPDFLKD--GEETVDWMSIYLSEENI----- 814
Db 808 NVPAQEEHGEERGVLEPTQQLAAAAVPMQKTVGQGEVDW---LDGEKVKBEQEV 864
QY 815 -----PNN---ADTSPFSESVG--SDAPSEAKYDLAEWYLEIGDRDAAAE-----TVQK 859
Db 865 FVHSGPNSQKAADVTYDSEVMGVAGCQEKESQVSLSEEGEMETDVEKEKRETKPEQV 924
QY 860 LEEAE 864
Db 925 SEBGE 929

RESULT 11

US-08-978-277A-2

; Sequence 2, Application US/08978277A

; Patent No. 6582956

; GENERAL INFORMATION:

; APPLICANT: Gelman, Irwin H.

; TITLE OF INVENTION: TUMOR SUPPRESSOR GENE

; NUMBER OF SEQUENCES: 20

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Brumbaugh, Graves, Donohue & Raymond

STREET: 30 Rockefeller Plaza
CITY: New York
STATE: NY
COUNTRY: USA
ZIP: 10112-0228
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASCSQ Version 1.5
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/978,277A
FILING DATE:
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/665,401
FILING DATE: 18-JUN-1996
ATTORNEY/AGENT INFORMATION:
NAME: Clark, Richard S
REGISTRATION NUMBER: 26,154
REFERENCE/DOCKET NUMBER: A30558 - 165/34008
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-408-2558
TELEFAX: 212-765-2519
TELEX:

INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1346 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: internal
ORIGINAL SOURCE:
US-08-978-277A-2

Query Match
Best Local Similarity 5.3%; Score 235; DB 4; Length 1346;
Matches 171; Conservative 102; Mismatches 268; Indels 304; Gaps 37;
QY 258 IKAEQPKQAKPAETAMPSEPKQATVEK-PVEKPAKVAEPAKAEKPAVREP-- 314
DB 151 IHTESPESADEQKGESSAPSEPEPTCLEKPLEAPRM-----GKLRKELLRGEK 204
QY 315 -----VPAANTAASETA---AESAPQ---EAAASAIPTDPTDETGNVSEPEVQSAEE 362
DB 205 KDHSIGILQKQDGTQETVRRPSESDEEKEKVKSAATLSSTSTVSEMDEKVTGGEOK 264
QY 363 TE-----SGLFGSYTLLLAGGGAALIALLLLR 393
DB 265 PREPKRVDTSVSWEALICVGSKKPKARKASSDINGPRTL---GGGQ-----S 310
QY 394 QSK-----RARRTEHSVEEPEPLDADGIEITFAETPATPPEA----- 436
DB 311 QSRGQOQRORRTDAPASTQEQDOA-----QGSSEPEPAGSPSEGEVSTWE 358
QY 437 -----PKNDVNDTLALDGESEELSAKQTFDVEDTP--SNRID--LDPDLAAQNG 485
DB 359 SFKRLVTPPKSKSL-----EEKEGRILVVGAGCPRLSRVKNLGFPLRNSPDG 411
QY 486 ILSGALTOEETQKRADADWNAIESTDSVYEPETFPNYPNVEIIVDTPPEPSVAQTAEN- 544
DB 412 GRKGQNGRQEQ-----TVEDSGPVEINEDEPDVPAVPLSEYD 450
QY 545 --KPEIVDTDFSDNLPS-----NNHI-----GTEETASAKPASPGLAGF 582
DB 451 AVEREKWEAQGNALPSCWGCVVSEELSKTLVHTVSAVIDGTRAVTSVEERSPSWISAS 510
QY 583 L-----KASSP--ETILEKTVAEVQTP-----EELHDFLKVYETDAVAE--TA 621
DB 511 VTEPLEHTAGEAMPPEVEVTEKDI IAEETPVLQTLPPEGKADHDDMTSEVDFTSEAVTA 570

QY 622 PETPD-----FNAADDLSALLQPAEAPSVEENITET-----VAETPD----- 659
DB 571 TETSEALRTBEVTEASGAETTDWMSAVSQLTSDPTTTEATPVQVEGGVLDTEEBRQ 630
QY 660 ----FNATADDLSALLQPSVPA-----VEENA----- 683
DB 631 TQAILQAVADKVK---BESQVPATQTVORTQSKALEKVEEVEEDSEVLASEKEDVMPKG 687
QY 684 -----AEIVAD-----DLSALLQPAEAPAEVNV 707
DB 688 PVQAGAGHLAGQSETCQATPESLEVPEVTADVDHVATCQVILQQLMEQAVAPESSETL 747
QY 708 TETVAETSDPHTAADDLSALLQPAEVPVAVENVTKTVAE----IDFNATADDLSALLOPS 764
DB 748 TDSETNGSTPLADSDTADGTQDETIDSQDSKATAAARQSQVTEEEAATAQKEEPTILPN 807
QY 765 EYPAVEENAAEITLET--PDSNTSEADALPDFLD--GEETVDWMSIVLSENI----- 814
DB 808 NYPAGEEHGEBFGDRDVLPTQOELAAAPVQWQTEVQGEVDM---LDGEKVKKEQEV 864
QY 815 -----PNN---ADTSFPSESVG--SDAPSEAKYDLAEMYLEIGDRDAAAE-----TVQKL 859
DB 865 FVHSGPNSQKAADVTYDSEVGVAGCQEKSETEVQSLSEBSEMETDVEKEKRETKPEQV 924
QY 860 LEEAP 864
DB 925 SEEGE 929

RESULT 12

US-09-134-001C-4463
Sequence 4463, Application US/09134001C
Patent No. 6380370
GENERAL INFORMATION:
APPLICANT: Lynn Doucette-Stamm et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
FILE REFERENCE: GTC-007
CURRENT APPLICATION NUMBER: US/09/134,001C
CURRENT FILING DATE: 1998-08-13
PRIOR APPLICATION NUMBER: US 60/064,964
PRIOR FILING DATE: 1997-11-08
PRIOR APPLICATION NUMBER: US 60/055,779
PRIOR FILING DATE: 1997-08-14
NUMBER OF SEQ ID NOS: 5674
SEQ ID NO 4463
LENGTH: 2137
TYPE: PRT
ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4463

Query Match
Best Local Similarity 5.3%; Score 235; DB 4; Length 2137;
Matches 154; Conservative 177; Mismatches 407; Indels 160; Gaps 26;
QY 11 PMMTKPTDCTRSNRIOPTPHRG-----YILKNRQIKLIAASVAVASQ 55
DB 750 PQGLKF-DASTNSIVGTPTQIGTNTITESTDASGNKTTTKINYEVRNRSASDSTST-- 806
QY 56 AHAGLGLNIQNLDEPEPSGJTVTGEEAKALLGGSVTVSEKGLTAKVHKLGDKAVIAV 115
DB 807 -----IVNSVSTISNSTLS-----DSVKASOSLSTSK--SLSISLAST 845
QY 116 SSEQAVRDPVLVFRICAGAQVREYTAILDPVGVSPKTSALSDGKTHRTAPTASQENQ 175
DB 846 SNSTSI-----QASESASTSKQLESASTSTSDSASARKSESTSKS----- 888
QY 176 NAKALRKTDKDSANAAVKPATNGKHTVTKGETVKQIAAAAPKHLTLEQVAD-ALLKA 234
DB 889 -TSLSESTSTSVSDSASVSTSEASTSTSVSGSTSTSTSDS-----TSTSTSDSASIKA 941
QY 235 NPNVSAHGLRAGSVLHLPNLNRKAEQPKQAKPAETAMPSEPEKQATVEKPEVKP 294


```

; Patent No. 6245337
; GENERAL INFORMATION:
; APPLICANT: St. Geme III, Joseph W.
; APPLICANT: Falkow, Stanley
; TITLE OF INVENTION: Haemophilus Adherence and Penetration
; TITLE OF INVENTION: Protein
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hobbach, Test, Albritton & Herbert
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: United States
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; FILING DATE: 25-AUG-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Trecartin, Richard F.
; REGISTRATION NUMBER: 31,801
; REFERENCE/DOCKET NUMBER: A-59941/RFT/RMS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1848 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
;
US-08-296-791-6

```

```

Query Match      5.3%; Score 232.5; DB 3; Length 1848;
Best Local Similarity 19.2%; Pred. No. 5.9e-08;
Matches 185; Conservative 140; Mismatches 377; Indels 261; Gaps 41;

QY 35 LKNNRQIKLIAASVAVASPOAHG-----LGGINI-QSNLDEPPFGSITVT 80
DB 766 LYSGRNVANITSNITASNAAQVHIGYKGTGTVCVRSYDTGYVTCHNSLSEKALNSFNP 825
QY 81 GEEAKA-LLGGGVTVSSEKGLTAKVKLGDKAVIAVSSEQAVRDPVLFRIGAGAQVREY 139
DB 826 NLRGVNUTENASFTLGRANLFGTIQSIGTSQVNLKNSH-----WHLTGNSVNNQL 877
QY 140 TAILDPVGYSPKTSALSDGKTHRTKAPTAESENQNAKALRK-----TDKK 186
DB 878 N-----LTNGHIHLNAQNDANKVTTYNTLVNSLSGNGSFYVWVDFTNK 922
QY 187 DSANAAVPAVNGKHTV--RKGETVKGIAAARPKH--LILEQVADA-----LLKANP 236
DB 923 SNKVVVNKSATGNFTLQVADKTGE-----PNHNELTLDASNATRNLEVTLANG 972
QY 237 NV-----SAHGRU-----RAGSVLHIPLNRIKAEQPKPTAK-----PKA 272
DB 973 SVDRGAWKYKLRNVGRDLYNPEVEKRNQVDTNITPTNDIQADAPSQNSNEIARV 1032
QY 273 ETASMPSPSKQATV--EKPVKPEAKVAAPAEKAPVAPRPPVPAANTAA-----322
DB 1033 ETPVPPPPAPATSAIAEQPETRP-AETAQPAEETNTANSTETAPKSDTATQNTENPSE 1091
QY 323 ---SET-----AASNPQEAASAIDTPTDETGNVASEVPEQVS 358
DB 1092 SVPSETTEKVAENPPQNETVAKNEQATEPTPQNGEVAKEDQPTVEANTQNTQATQSEG 1151
QY 359 ABEETESGLFGSYLLLAGGAALIALILLRLAQSKRRARTEESVPEEPD--LDDAA 416
DB 1152 KTEETQT-----AETKSEPTSVTSVENQPEKTSQST 1184

```

```

QY 417 DDGIEI---TFAEVET-----PATPEAPKNDVNDTLALDGSBEELSAKQ 459
DB 1185 EDKVVVEKEKAKVETEETOKAPQVTSKEPKQAPAEPEEPTDTNAAEQALQQTPTT 1244
QY 460 TFDVETDTPSNRDLDFDLSAAQNG-----ILSGALTQDEETQKADADNNAIESIDS 513
DB 1245 VAAAEITSPNSKPAEETQOPSEKTNABFPVTPVVSSENTATQPTETEETAKVE---KEKTQE 1301
QY 514 VYBPETENPNVPIVIDTPEPSVAQTAENKPTVDTDFSDNLPNSNNHIGTEETASAKP 573
DB 1302 V--PQVASQSPKQ---EQPAKPAQAT---KPOAEP--RENVLTTKVG--EQPQOAP 1350
QY 574 ASPS---GLAGFLKASSPETILEKTVAEVQTPBELHDFLKYVETDAVAETAPETPDFNAA 630
DB 1351 QTQSTAVPTTGETAANSKPAKPAQAKPQTPAPARENVTVTNKEPQSTSATVSTEQPA 1410
QY 631 ADDLSALLOPAEAPSVEENITETVAETPDFNATADDLSALLOPSEVPAVENAAEIVADD 690
DB 1411 KETSSNVEQAPENSINTGSAATTMTET-----AEKSKPKQME---TVTEND 1453
QY 691 LSALLOPAEAPAVEENVETETVAETSDFHATAADDLSALLOPSEVPAVENVTKTVAEIPDF 750
DB 1454 ---RQP-EANTVADNSVANNSESE--SKSRRRSYSQPKETSAEETTVASTQ-----1500
QY 751 NATADDLSALLOPSE---VPAVEENAAEITILETPDSNTSEADALPDFLXGEEETVDSI 807
DB 1501 ETTVDNSVSTPKPRSRRTSRSVQTNSE--PVELPTENAENAENVQ-----1544
QY 808 YLSEENTPNAD-----TSFPEBSVGSDAPSEAKYDLAEMYLEIGDRDAAAEVTKLLEE 862
DB 1545 --SGNNVANSQPALRNLTSKNTNAVISNAMAQAF---VALNVG---KAVSQHSQLEMMN 1596
QY 863 AEG 865
DB 1597 NEG 1599

```

```

RESULT 15
US-09-839-996-6
; Sequence 6, Application US/09839996
; Patent No. 6642371
; GENERAL INFORMATION:
; APPLICANT: St. Geme III, Joseph W.
; APPLICANT: Falkow, Stanley
; TITLE OF INVENTION: Haemophilus Adherence and Penetration
; TITLE OF INVENTION: Protein
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hobbach, Test, Albritton & Herbert
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: United States
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/839,996
; FILING DATE: 20-Apr-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/296,791
; FILING DATE: 25-AUG-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Trecartin, Richard F.
; REGISTRATION NUMBER: 31,801
; REFERENCE/DOCKET NUMBER: A-59941/RFT/RMS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989

```

; TELEFAX: (415) 398-3249

; TELEX: 910 277299

; INFORMATION FOR SEQ ID NO: 6:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1848 amino acids

; TYPE: amino acid

; TOPOLOGY: unknown

; SEQUENCE DESCRIPTION: SEQ ID NO: 6:

US-09-839-996-6

Query Match 5.3%; Score 232.5; DB 4; Length 1848;

Best Local Similarity 19.2%; Pred. No. 5.9e-08;

Matches 185; Conservative 140; Mismatches 377; Indels 261; Gaps 41;

```
QY 35 LKNROIKLIAAVAVAAPOAHAG-----LGGLNI-QSNLDEPFGSITVT 80
Db 766 LYSGRVANITSNITASNNAQVHIGYKGTVCVRSDYTGTVYVCHNSNLSEKALNSFNPT 825
QY 81 GEBAKA-LLGGGSVTVSEKGLTAKVHKLGDKAVIAVSEQAVRDPVLVFRIGAGAVREY 139
Db 826 NLRGNVNLTENASFTLGKANLFGTIQSIGTSQVNLKENS-----WHLTGNSNVNQL 877
QY 140 TAILDPVGYSPKTSALSQKTHKTAFTAESQENQNAKALRK-----TDKK 186
Db 878 N-----LTNGHIHLNAQNDANKVTTYNTLTVNSLSGNGSFYVWVDFTNK 922
QY 187 DSANAAVKPAYNGKTHV--RKGETVKQIAAAIRPKH--LTLEQVADA-----LLKANP 236
Db 923 SNKVVNKSATGNTLQVADKGE-----PNNELTLFDASNATRNNLVLIANG 972
QY 237 NV-----SAHGRL-----RAGSVLHPIPNLRKAEQPKQTAK---PKA 272
Db 973 SVDRGAWKYKLRNVNGRYDLNPEVEKRNQTVDTNITTENDIQADAPSAQSNEETARV 1032
QY 273 ETASMPSEPSKQATV--EKPEVPEAKVAPEAKPAKPAVRPEPVAANTAA-----322
Db 1033 ETPVPPAPATESAIASEQPETRP-AETAQPAEMEETNTANSTETAPKSDTATQTEPNSE 1091
QY 323 ---SET-----AESAPOEAAASALDPTDGTGNAVSEPVPEQVS 358
Db 1092 SVPESETTEKVAENPPQNEETVAKNEQATEPTQNGEVAKEDQPTVEANTQTNEATQSEG 1151
QY 359 ABETESGLFGGSYTLILLAGGGAALIALLLLRLLAKSKARRTEESVPEEEDP--LDDAA 416
Db 1152 KTEETQT-----AETKSEPTSVTVSENOPEKTVSQST 1184
QY 417 DQGIET---TFARVET-----PATPEPAPKVDVNDTLALDGESEELSAKQ 459
Db 1185 EDKVVVEKEKAKVETEETQKAPQVTSKEPPKQAPAEPEVPTDTNAEEAALQQTQPTT 1244
QY 460 TFDVETDTPSNRIDLDFDSLAAQNG-----ILSGALTQDEETOKRADADWNAIESTDS 513
Db 1245 VAAAEITSPNSKPAETQOPSEKTAEPVTPVVSSENTATQPTETEETAKVE---KEKTQE 1301
QY 514 VYEPETFPNVPVEIIVIDTPESVQAQTAENKPEVTDTPFDNLPSNNHIGTEETASAKP 573
Db 1302 V--PQVSAQSPKQ---EQPAKPAQCT--KQPAEPA--RENVLTTKNVG-EPQPAQAP 1350
QY 574 ASDP---GLAGFLKASPELTKVAVQVTPPELHDFLKVETDVAETAETAPETPDFNAA 630
Db 1351 QTQSTAVPTTGETAANSKPAKPAQAKPQATEPARENVSTVNTKEPQSQTSATVSTEQPA 1410
QY 631 ADDLSALLQPAEAPSVENITETVAETPDFNATADDLSALLQSEVPAPVEENAAEIVADD 690
Db 1411 KETSSNVEQAPENSINTGSAITWTET-----AEKSDKQME---TVTEND 1453
QY 691 LSALLQPAEAPAVEENVETVAETSDPHTAADDLSALLQPAEPAVEENVTKTVAEIPDF 750
Db 1454 ---RQP-EANTVADNSVANNSESSE--SKSRRRRSVSPKETSARETTVASTQ-----1500
QY 751 NATADDLSALLQPE---VPAVEENAAEITLETDPDNTSEADALPDFLKDGBEETVDSI 807
Db 1501 ETTVDNSVSTPKPRSRRTSRVQTSNYSY-EVELPTENAENVQ-----1544
```

```
QY 808 YLSEENIPNNAD-----TSFPSESVGSDAPSEAKYDILAEMYLEIGDRDAAAETVOKLEE 862
Db 1545 --SGNNVANSGPALRNLTSKNTNAVVISNMAKAQF-----VALNVG--KAVSQHISQLEMN 1596
QY 863 AEG 865
Db 1597 NEG 1599
```

Search completed: September 2, 2004, 19:21:56
Job time : 27 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 2, 2004, 19:20:52 ; Search time 144 Seconds
(without alignments)
1925.398 Million cell updates/sec

Title: US-09-743-674-2

Perfect score: 4404

Sequence: 1 MPAGRLPRCPMMTKFTDCT.....EEAEGDVLKRAQALAEELGI 880

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1298764 seqs, 315065143 residues

Total number of hits satisfying chosen parameters: 1298764

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications AA:*
1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep:*
10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep:*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep:*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep:*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3434	78.0	801	12	US-10-282-122A-65373
2	2391.5	54.3	503	12	US-10-282-122A-65571
3	381.5	8.7	919	12	US-10-282-122A-66451
4	358	8.1	1095	14	US-10-128-714-8305
5	315.5	7.2	1786	9	US-09-742-096-3
6	315.5	7.2	1787	12	US-10-415-253-2
7	315	7.2	1000	14	US-10-128-714-3305
8	276.5	6.3	1772	16	US-10-437-963-143280
9	276.5	6.3	1844	16	US-10-437-963-143282
10	269	6.1	540	9	US-09-742-096-22
11	269	6.1	2364	12	US-10-205-331-66
12	260.5	5.9	1444	16	US-10-437-963-150334
13	258	5.9	1859	15	US-10-369-493-19525
14	253.5	5.8	3507	15	US-10-369-493-5784
15	253	5.7	1367	9	US-09-801-368-108

16	251	5.7	1348	12	US-10-282-122A-56877
17	250.5	5.7	630	9	US-09-742-096-5
18	244	5.5	1616	10	US-09-820-843A-16
19	244	5.5	1616	12	US-10-282-122A-63593
20	242.5	5.5	530	16	US-10-437-963-117065
21	238	5.4	2665	9	US-09-864-761-34248
22	238	5.4	3664	12	US-10-263-929-143
23	238	5.4	3664	14	US-10-177-293-423
24	238	5.4	3664	16	US-10-408-765A-2287
25	236	5.4	1596	9	US-09-902-432-4
26	236	5.4	2478	9	US-09-815-242-5816
27	236	5.4	2478	9	US-09-815-242-12967
28	236	5.4	6642	15	US-10-369-493-5013
29	235	5.3	1346	9	US-09-902-432-2
30	233	5.3	740	15	US-10-369-493-3700
31	233	5.3	1665	12	US-10-282-122A-71690
32	232.5	5.3	1848	10	US-09-839-996-6
33	232.5	5.3	1848	12	US-10-645-655-6
34	232.5	5.3	1848	14	US-10-080-505-6
35	232.5	5.3	1848	16	US-10-687-046-6
36	230	5.2	2435	12	US-10-282-122A-47453
37	229.5	5.2	2481	12	US-10-282-122A-43762
38	228.5	5.2	2368	9	US-09-815-242-5635
39	228.5	5.2	2368	9	US-09-815-242-12389
40	227.5	5.2	1831	12	US-10-282-122A-71033
41	227.5	5.2	2472	9	US-09-815-242-5064
42	226.5	5.1	1178	14	US-10-128-714-8240
43	226.5	5.1	1879	9	US-09-971-536-70
44	226.5	5.1	2402	16	US-10-661-809-20
45	226	5.1	570	12	US-10-415-253-4

ALIGNMENTS

RESULT 1

US-10-282-122A-65373
; Sequence 65373, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636


```
QY 435 PAKKNDVNDTLALDGESEBELSAKOTFOVETDTPSNRIDLDLDFDSLAAAQNGILSGALTD 494
Db 121 PAKKNDVNDTLALDGESEBELSAKOTFOVETDTPSNRIDLDLDFDSLAAAQNGILSGALTD 180
QY 495 EBTQKRADADNNAIESTDSVYEPFNPVNPVIEVIDTPEPESVAQTAENKPEVDVDTDFS 554
Db 181 EBTQKRADADNNAIESTDSVYEPFNPVNPVIEVIDTPEPESVAQTAENKPEVDVDTDFS 240
QY 555 DMLPNNHIGTETASAKPASPSGLAGFLKASSPETILEKTVAEVOPTPELHDFLKVYET 614
Db 241 DMLPNNHIGTETASAKPASPSGLAGFLKASSPETILEKTVAEVOPTPELHDFLKVYET 300
QY 615 DAVAETAPETPDFAADLDSALLOPAEAPSVENITETVAETPDNATADDLSALLQPS 674
Db 301 DAVAETAPETPDFAADLDSALLOPAEAPSVENITETVAETPDNATADDLSALLQPS 360
QY 675 EYPVAVEENAAETVADLDSALLOPAEAPAEVENVETVAETSDFTAAADLSALLQPAEVP 734
Db 361 EYPVAVEENAAETVADLDSALLOPAEAPAEVENVETVAETSDFTAAADLSALLQPAEVP 420
QY 735 AVEENVTKVAEIPDNATADDLSALLQPAEAPAEVENVETVAETSDFTAAADLSALLQPAEVP 794
Db 421 AVEENAAETV-----ADLSALLQPAEAPAEVENVETVAETSDFTAAADLSALLQPAEVP 471
QY 795 LKDGEEETVDWSIYLSEENIPNNAATSF 822
Db 472 LKDGEEETVDWSIYLSEENIPNNAATSF 499

RESULT 3
US-10-282-122A-66451
; Sequence 66451, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Cart, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
```

```
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 66451
; LENGTH: 919
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-10-282-122A-66451

Query Match      8.7%; Score 381.5; DB 12; Length 919;
Best Local Similarity 23.0%; Pred No. 2.7e-15;
Matches 232; Conservative 134; Mismatches 362; Indels 281; Gaps 43;

QY 41 IKLIAASVAVAAASFOAHA-GLGGLNIQSNLDPPGSGSIIVTGEBAKALGGGSVTVS--- 96
   ::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 8 VRAIAAA-SVLTSMAHGLGELITLKSALNPDLDAEIELL--EVRD-LGSGEVIPSLAS 63

QY 97 -BKGITAKVHKI-----GDKAVTAVSSEQAVRDPVLVFRIGA---GAQVR 137
   |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 64 PEEFSKAGVDRLYLTDLKFTPVVKNPGKSVIRVTSKPVQEPYLNFLVQVLPNGRLLR 123

QY 138 EYTAILDPVGYSPKTKSALSDBGKTHRKTAAPTAESENQNAKALRTDKKDSANAAVK--- 194
   |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 124 EYVILLDPPLYSFOAAASAPQPV---SAPRATG-----APRAPQAPAPVRTTA 169

QY 195 PAYNGKTHTVRKGETVTKQIAAAIRPKHLTLEQVADALKANPNVSAHG---RLRAGSVLH 251
   |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 170 PAGSDTYRTV-SNDTLWEIAQRNRTDRVSVPOAMLAFLQELNPGAFVDGNINRLKSGQVLR 228

QY 252 IPNLNRIKAEQPKQTKAKPAETAS-----MPSEPSK-QAT 286
   |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 229 IPTEQMLERSPREALSQVQAQNSQWRGSRNPAGSAGARQLDATQRNAAGSAPSKVDAT 288

QY 287 ----- 286

Db 289 DNLRLVSGEGKASKGADKGGKDSKAIADTLAVTKESLDSTRENEELQSRMQDLOSQLD 348

QY 287 -VEKPEKEPEAKVA-----APEAKAEKPAVRPFVPAANTA 321
   ::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 349 KLQKLIQLKDAQLAKLQGLGAEQGAQPNAAALPDASQPNAAAQAQAPGTPAAAAPT 408

QY 322 ASETAESAEPQEAASADITPTDE--TGNVSEPVQVSAEETESGLFGGSYTLLAG-- 378
   |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 409 APAGEAPAPAPQFPVAPPAPAAEKPPAPAPAPAPVQAAEQAPAFSL-----DELLANP 463

QY 379 -----GGAALIALLLLRLAQSKARTEESVP-----EEPDLDAAADGI-EITF 424
   |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 464 LMLAVIGGALLALVLVLMILSRNAQKEKEAQAFAADTGEQEDALDLGKGFDDLL 523

QY 425 AEVETPATPEPAPKNDVNDTLALDGESEBEL-----SAKQTFDVTDTTPSNRIDLD 475
   |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 524 DEPE-POVAAPVAPQVEKTTAQTSDALGEADIVAYGRFNQAELLQNAIYDEP-QRTDLR 581

QY 476 FDSLAA-AQNGILSGALTQDEETQKRADADNNAIESTDSVYEPFNPVNPVIEVIDTPE 534
   :|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 582 LKLMVEYAEIMGDRGFAQENELREIGGAQ-POVEQLKSR-----PAMVAVAVA 631

QY 535 PESVAQTAENKPEVDVDTDFSNDLPNNHIGTETASAKP-ASPSGLAGFLKASSPETILE 593
   :|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 632 GLAGAKLAQDELDSFSLD-DLSLDSG-----AAKPDAAQDLDDAFDLSLDDLGD 683

QY 594 KTVAEVQTPPE-ELHDFLKVYETDAVAETAPETPDFAADLSALLQPAEAPSVENITE 652
   |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 684 DVQADLKSDSGALDDLTLDSLDLDAASTPADKP-----VDDLDFGLDFAE----- 728

QY 653 TVAETPDFAATADDLSALLQPSVPAVEENAAEIVADDLSALLQ-----PRAEPAVENVTE 709
   |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 729 -LAETPS-QPKHDDLGDGFSLDLADP--EDKLSD---DDFLSLINDVPPAAPADNEFTLD 781

QY 710 TVAETSDFHTAAD--DLSALLQPAEVPVAVENVTKVAEIPDNATADDLSALLQSEVP 767
   |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 782 TEAAEEPALSLPDDFSLADEPTPEAPAEKEDGSFAAQLDVEVSAQLDELAS----- 833

QY 768 AVEENAAETVLETPDSNT---SEADALPDLKDGEEETVDWSIYLSEENIPNNAATSFPS 824
   |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
```

Db 834 -----NLDPEKATPSPSARDAAVASALDGD-----ADDDPDF 866

Qy 825 ESVGSDAPSEAKYDLAEMYLEIGDRDRAAAETVQKLLAEAGDVLKRAQA 873

Db 867 LSGADEAAT--KLDLARAYIDMGSEGDILDEVL--AEGNDSQQA 911

RESULT 4

US-10-128-714-8305

; Sequence 8305, Application US/10128714

; Publication No. US20030119013A1

; GENERAL INFORMATION:

; APPLICANT: Jiang, Bo

; APPLICANT: Hu, Wengqi

; APPLICANT: Tishkoff, Daniel

; APPLICANT: Zamudio, Carlos

; APPLICANT: Eroskin, Alexey M

; APPLICANT: Lenieux, Sebastien M

; TITLE OF INVENTION: Identification of Essential Genes in *Aspergillus fumigatus* and

; TITLE OF INVENTION: Methods of Use

; FILE REFERENCE: 10182-018-999

; CURRENT APPLICATION NUMBER: US/10/128,714

; CURRENT FILING DATE: 2002-04-23

; PRIOR APPLICATION NUMBER: US 60/285,697

; PRIOR FILING DATE: 2001-04-23

; PRIOR APPLICATION NUMBER: US 60/287,066

; PRIOR FILING DATE: 2001-04-27

; PRIOR APPLICATION NUMBER: US 60/295,890

; PRIOR FILING DATE: 2001-06-05

; PRIOR APPLICATION NUMBER: US 60/303,899

; PRIOR FILING DATE: 2001-07-09

; PRIOR APPLICATION NUMBER: US 60/316,362

; PRIOR FILING DATE: 2001-08-31

; NUMBER OF SEQ ID NOS: 8603

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 8305

; LENGTH: 1095

; TYPE: PRT

; ORGANISM: *Aspergillus fumigatus*

; FEATURE:

; NAME/KEY: MISC FEATURE

; LOCATION: (1095)..(1095)

; OTHER INFORMATION: X= any amino acid

US-10-128-714-8305

Query Match 8.1%; Score 358; DB 14; Length 1095;

Best Local Similarity 22.1%; Pred. No. 1e-13;

Matches 233; Conservative 140; Mismatches 412; Indels 270; Gaps 44;

Qy 38 NRQIKLIAASVAVASFOAHAGLGLNQTQSLNDEPFGSGITVTGEEAKALGGGSGVTVSE 97

Db 91 NNILTVGSSSIAAAPGAHA-----VEKETETPLENGASEKVEETAKEPGVQSTTTTTE 144

Qy 98 KGLTAKVHKLGDKAVIA-----VSSQAVRD--PVLVFRIGAGAVRE--YTALIDPVGYS 149

Db 145 AVKDEVEKTDSDVVVEKDLHVESEPAQVEKPKLVSEPKSPDIHEDAEQPSANAAE 204

Qy 150 PKTKSALS-----DGKTHRKTAFTAESQENQAKARKTDKOSANAAVKPAYNGK 200

Db 205 PKTNGTHSSEQISGQPDVAATEKVEPVEEKAADLTTKTKVANKPBEIPSTLSAEKPE 264

Qy 201 THTVRKGE-----TVKQIAAIRPKHL-----TLEQVADALLKANPNVSAHGLRL 245

Db 265 TKVEKEADQKLDLEIVPATVEKTEAKSEVPSESTLEKSEQEAPAAQ--TVTTEEPLV 323

Qy 246 AGSVLHLPNLRIRKAEQ--PKQTA--KPKAETAS---MPSEP---SKQATV----- 287

Db 324 DSKPVEISSABSTAEKAKPTMTTDEAPLAKSEKTVPAEPIAEKQVAILIGDEPLSRN 383

Qy 288 ---EKVKEKP--EAKVAPEAKAEKPAVRPEPVAANTAASETAESAFAPEAAASAI--DTP 342

Db 384 NVEEQPKQKSVKAEAKADEAVLEEP-----IKESAPEEIPEDSRNAVADAP 430

Qy 343 TDETG-----NAVSEPVQVSAREETESGLFGGSYTLLLAGGGAALIALLLRL 392

Db 431 VTSESTTEKVEVTAPDAEKEPALKESAKEPIQEIBPEASEKEVAE-----TPV 479

Qy 393 AQSKRAARTESV--PREEPDLDDAADDGI-----EITFA-EVETPATPEPAP 437

Db 480 TESSTTEKVEVTALDAEKEPALKEESTKEPIHEBQTIETVTVTAETAAKEPQAQATEEPPVA 539

Qy 438 KNDVNDTILALDGESEE---ELSAKQTFDVT-----DTPSNRIDLDFDSLAAQNGI 486

Db 540 NESINK--EVEATTEETAGQESAKEPISNETAIRVTAQADESTKELSGTF---AAAEIT 594

Qy 487 LSGALTQDEET-----OKRADADWNAIESTD----- 512

Db 595 GKGAETHTEVSTAAGPAKEFVNDASAVKSAPTEVTQBTPTKDDSAQVVVVGKEAPEQIAEE 654

Qy 513 -SVYEPETFNPNPVEIVIDTPEP-----ESVAQTAENKPEPTVD 550

Db 655 AAETSAANPEATEETSTTAEBPAKEPGESETPCQILPAAEPKEATAEBAAKEESTIE 714

Qy 551 TDFSDNL-----PSNNHIGTEETASAKPASPSGLAGFLKASSPETTILEKTVAEVQTPPE 604

Db 715 T-VSEPLEADVKEPAQDDPATETLTEDKPVTOESVAE--AAPKPSVSEPVVSEGTAKTS 771

Qy 605 LHDFLKVYETDAVA-----ETAPETPPFNAAD--DLSALLQPAEA 643

Db 772 AVDEAESSEKPAVVTTETKEPGEFEBHPVTRELAKSESEASDKAAKSGEAPAAPETILEA 831

Qy 644 PSVEENITETVAETPDFNATADDLSALLQPSVPAVEENAAEIV-----ADDSALLQOP 697

Db 832 PAPETVPEPVSESPETEVAKE-----PATVESTQETPTTGTGEQAKGEVAINEP 881

Qy 698 AEAPAVEENVTTVAETSDFTAAADLSALLQPAEVPVAVENVTKTVAEIPDFNATADDL 757

Db 882 TEATATEEAVPEEVSEVKG--EIEEPVAAVKKSELPADEPLCNDGCS--PEEITTEEA 936

Qy 758 SALLQPSVPAVEENAAEITLETDPDSNTSEADALP-----DFLKDGE-----EETVDWS 806

Db 937 VQKTKASDEPVEETNATEVTIKEPTA--TETTESAPVKEATETVKEPEAVAPVABEPVKTA 995

Qy 807 I-----YLSEENI-----PNNADTSFPSESVGSDAPSEAKYDLA-----EMVL 844

Db 996 IAGDNTTEKLLAEETVSKDVTEPVSVAEKVVPSEPAETEEPVSEPTSEATERAEPEASI 1055

Qy 845 EIGDRDAAAE--TVOKLLEEAEGDVLKRAQALAE 877

Db 1056 QVPAQDEVADVAAIEEQEQAQVAEPVTKPAEPAAE 1090

RESULT 5

US-09-742-096-3

; Sequence 3, Application US/09742096

; Patent No. US20020155441A1

; GENERAL INFORMATION:

; APPLICANT: DRUILHE, PIERRE

; APPLICANT: DAUBERES, PIERRE

; TITLE OF INVENTION: MALARIAL PRE-ERYTHROCYTIC STAGE POLYPEPTIDE MOLECULES

; FILE REFERENCE: 200773USODIV

; CURRENT APPLICATION NUMBER: US/09/742,096

; CURRENT FILING DATE: 2000-12-22

; PRIOR APPLICATION NUMBER: US 08/973,642

; PRIOR FILING DATE: 1998-02-06

; PRIOR APPLICATION NUMBER: PCT/FR96/00894

; PRIOR FILING DATE: 1996-06-12

; PRIOR APPLICATION NUMBER: FR 95/07007

; NUMBER OF SEQ ID NOS: 29

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 3

; LENGTH: 1786

; TYPE: PRT

; ORGANISM: *P. falciparum*

US-09-742-096-3

```
Query Match          7.2%; Score 315.5; DB 9; Length 1786;
Best Local Similarity 20.2%; Pred. No. 9.5e-11;
Matches 167; Conservative 147; Mismatches 331; Indels 183; Gaps 26;

QY 64 NIOSNLDPEPSSGIVTGBEAKALGGSVTVSEKGLTAKVHKLGDKAV---IAVSSEQA 120
DB 242 NVEENVDGSGVASSVEES-----IASSVDESIDSSTEENVAPTVEEI 286
QY 121 VRDPVLVFRIGAGAQVREYTAILDPVGYSPKTKSALSDBGKTHKRTAPTAE--QENQNAK 178
DB 287 VAPSVV-----ESVAPSVESVEEENVEES 310
QY 179 ALRKTDKDSANAAVKPAYNGKTHTVRKGE-----TVKQIAAAIRPKHLTLEQ-VADALL 232
DB 311 VAENVEESVAENVEESVAENVEESVAENVEEIVAPTVEEIVAP-----TVEEIVAPSVV 364
QY 233 KA-NPNVSAHGRLRAGSVLHINLNRIKABQPKQAKPAETASMPSEPSKQATVEKPV 291
DB 365 ESVAPSVESVE-----ENVEESVAENVEESVAENVEESVAENVEESVAENVEESV 415
QY 292 EKPEAKVAPE-----AKAEKPAVRPEPVPAAANTAASETAESAPOEAAASAITDPTDET 346
DB 416 AENVEEIVAPTVEEIVAPTVEEIVAPSVVESVAPSVESVEEENVEESVAENVEESVAENV 475
QY 347 GNAVSPFVEQVSAEETESGLFGGSYTLLLAGGGAALIALLLRLAQSRRARTTESVP 406
DB 476 EESVAENVEESVAENVEES-----VAENVEESVA 504
QY 407 EE-EPDLDDAADDGIITFAEVETPATPPAPKNDVNDTLALDGESEELSAKQTFDDET 465
DB 505 ENVEESVAENVEEIVAPTVEEIVAPTVEEIVAPSVVESVAPSVESVEE-----NVEE 557
QY 466 DTPSNRIDLDFSLAAQNGILSGALTQDBEETQKRADADWNALESSTDVVEPETFNPYP 525
DB 558 SVAEN-----VEESVAENVEESVAENVEESVAENVEEIVAPTVEEIVAP 601
QY 526 -VEIVIDTPEPESVAOTAENK-PETVDTDFSDNLPNNHIGTETASAKPASPGLAGFL 583
DB 602 TVEEIVAPSVVESVAPSVESVEEENVEESVAENVEESVAENVEESV-----VAPSVEE 647
QY 584 KASSPTILEKTVAEVQTPPEELHDFLKVYETDAVETAPETPDFNAAADLSALLQPAEA 643
DB 648 -AENVEEIVAPTVEEIVAP-TVEEIVAPSVVESVAPSVESVEEEN-VEESVAENVEESVA 704
QY 644 PSVEENITETVAPTDFNATADLSALLQPSVPVPAVEENAAEIVADLSDALLQPAEAPAV 703
DB 705 ENVEESVAENVEE-----SVAENVEEIVAPTVEEIVAPTVEEIVAPSVVESVAPSVESV 759
QY 704 EENVETVAETSDFHTAADDLSALLQPAEVPVAVEENVTKTVAEIPDFNATADLSALLQ 763
DB 760 EENVEESVAENVE-----ESVAENVEESVAENVEESVAPTVEEI-----VAPSVEE 805
QY 764 SEVPVAEENAAEITLETDPDSNTSEADALPDFLKDGEETVDMWSIYISEENIPNNAUTSFP 823
DB 806 SVAPSVESVA-----ENVATNLSD-NLLSNLLGGIETEIEIKDSILNEIBEVENVTVTI- 859
QY 824 SESVGSDAPESEAKYDLAEMYLEIGDRDAAAEIVQKLLLEAEAGDVLKRA 871
DB 860 LENV-BETTAESVTTFSNILEEIQENTITNDTIEEKLLEHENVLSAA 906

RESULT 6
US-10-415-253-2
; Sequence 2, Application US/10415253
; Publication No. US20040067236A1
; GENERAL INFORMATION:
; APPLICANT: Cohen, Joe
; APPLICANT: Druilhe, Pierre
; TITLE OF INVENTION: Immunogenic Compositions Comprising
; TITLE OF INVENTION: Liver Stage Malarial Antigens
; FILE REFERENCE: B45250
; CURRENT APPLICATION NUMBER: US/10/415,253
```

```
; CURRENT FILING DATE: 2003-04-25
; PRIOR APPLICATION NUMBER: PCT/EP01/12349
; PRIOR FILING DATE: 2001-10-23
; PRIOR APPLICATION NUMBER: EP00203724.0
; PRIOR FILING DATE: 2000-10-25
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 1787
; TYPE: PRT
; ORGANISM: K1 Parasite Clone
; US-10-415-253-2

Query Match          7.2%; Score 315.5; DB 12; Length 1787;
Best Local Similarity 20.2%; Pred. No. 9.5e-11;
Matches 167; Conservative 147; Mismatches 331; Indels 183; Gaps 26;

QY 64 NIOSNLDPEPSSGIVTGBEAKALGGSVTVSEKGLTAKVHKLGDKAV---IAVSSEQA 120
DB 242 NVEENVDGSGVASSVEES-----IASSVDESIDSSTEENVAPTVEEI 286
QY 121 VRDPVLVFRIGAGAQVREYTAILDPVGYSPKTKSALSDBGKTHKRTAPTAE--QENQNAK 178
DB 287 VAPSVV-----ESVAPSVESVEEENVEES 310
QY 179 ALRKTDKDSANAAVKPAYNGKTHTVRKGE-----TVKQIAAAIRPKHLTLEQ-VADALL 232
DB 311 VAENVEESVAENVEESVAENVEESVAENVEEIVAPTVEEIVAP-----TVEEIVAPSVV 364
QY 233 KA-NPNVSAHGRLRAGSVLHINLNRIKABQPKQAKPAETASMPSEPSKQATVEKPV 291
DB 365 ESVAPSVESVE-----ENVEESVAENVEESVAENVEESVAENVEESVAENVEESV 415
QY 292 EKPEAKVAPE-----AKAEKPAVRPEPVPAAANTAASETAESAPOEAAASAITDPTDET 346
DB 416 AENVEEIVAPTVEEIVAPTVEEIVAPSVVESVAPSVESVEEENVEESVAENVEESVAENV 475
QY 347 GNAVSPFVEQVSAEETESGLFGGSYTLLLAGGGAALIALLLRLAQSRRARTTESVP 406
DB 476 EESVAENVEESVAENVEES-----VAENVEESVA 504
QY 407 EE-EPDLDDAADDGIITFAEVETPATPPAPKNDVNDTLALDGESEELSAKQTFDDET 465
DB 505 ENVEESVAENVEEIVAPTVEEIVAPTVEEIVAPSVVESVAPSVESVEE-----NVEE 557
QY 466 DTPSNRIDLDFSLAAQNGILSGALTQDBEETQKRADADWNALESSTDVVEPETFNPYP 525
DB 558 SVAEN-----VEESVAENVEESVAENVEESVAENVEEIVAPTVEEIVAP 601
QY 526 -VEIVIDTPEPESVAOTAENK-PETVDTDFSDNLPNNHIGTETASAKPASPGLAGFL 583
DB 602 TVEEIVAPSVVESVAPSVESVEEENVEESVAENVEESVAENVEESV-----VAPSVEE 647
QY 584 KASSPTILEKTVAEVQTPPEELHDFLKVYETDAVETAPETPDFNAAADLSALLQPAEA 643
DB 648 -AENVEEIVAPTVEEIVAP-TVEEIVAPSVVESVAPSVESVEEEN-VEESVAENVEESVA 704
QY 644 PSVEENITETVAPTDFNATADLSALLQPSVPVPAVEENAAEIVADLSDALLQPAEAPAV 703
DB 705 ENVEESVAENVEE-----SVAENVEEIVAPTVEEIVAPTVEEIVAPSVVESVAPSVESV 759
QY 704 EENVETVAETSDFHTAADDLSALLQPAEVPVAVEENVTKTVAEIPDFNATADLSALLQ 763
DB 760 EENVEESVAENVE-----ESVAENVEESVAENVEESVAPTVEEI-----VAPSVEE 805
QY 764 SEVPVAEENAAEITLETDPDSNTSEADALPDFLKDGEETVDMWSIYISEENIPNNAUTSFP 823
DB 806 SVAPSVESVA-----ENVATNLSD-NLLSNLLGGIETEIEIKDSILNEIBEVENVTVTI- 859
QY 824 SESVGSDAPESEAKYDLAEMYLEIGDRDAAAEIVQKLLLEAEAGDVLKRA 871
DB 860 LENV-BETTAESVTTFSNILEEIQENTITNDTIEEKLLEHENVLSAA 906
```

```

Db      416 -ARTESAABESAPAEATEISISVGPEAAADAREGL-----REBSAHEEPAPVED--- 464
Qy      582 FLKASSPETILEKTVAE---VOTPEELHDFLKVYETDVAETAP---ETPDFNAAADOLS 635
Db      465 -APAEPAPVVEEPAAEESPSVEDPAPVVEEPAPAEADPAAAABEIPTEEPAPAEVPAEPA 523
Qy      636 ALLQPARAPS-VVEENITETVA-----ETPDFNATADDLSALLQPSVEVPVAVEENAAEIVADD 690
Db      524 PVQEPAEPTFPVKSPVEESAPAEERIPVEEAPAEEST-----PAEPAPVEVPAEI----- 575
Qy      691 L$ALLQPA--EAPAVEE-NVTEIV-AETSDFHAAADLSALLQPAEV-PA-VVEENVTKT 744
Db      576 -SAPVPEAPVEEPAEABEAPTEESARBEPAEMEEAPTEESASVSPASVEPAPVPGPVVEEP 634
Qy      745 AEIPDFNATADDLSALLQPSVEVPVAVEENAAEITLETPDFNSNTSADALPDFLKDGESEETVD 804
Db      635 A-----AAAEI-----PAEEFVVVEGAPVEESVPVKEAPVDA-----EFSSETPD 677
Qy      805 WSIYLSSEENIPNNAADTSFPSES 826
Db      678 PT---PVVEISRDIDALEPLEA 696

RESULT 8
US-10-437-963-143280
; Sequence 143280, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associat
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 143280
; LENGTH: 1772
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_44203C.1.1.pep
US-10-437-963-143280

```

Query Match	6.3%	Score 276.5;	DB 16;	Length 1772;
Best Local Similarity	21.2%;	Pred. No. 2.7e-08;		
Matches 212;	Conservative 142;	Mismatches 379;	Indels 267;	Gaps 42;
Qy	6	LPRCPMMTFTDCTRSNRITQPPTHRYIILKNRRQ-----IKLIAASVAASAQFAHAGL	60	
		: : : : : :	:	:
Db	481	VPEHSGITEKPSQTSADDVPKESTAAVVKQWQRNIGFQKPQQSVTDAKTPESR-GM	539	
		: : : : : :	:	:
Qy	61	GGLNIQSLNDPEPSSGSIWTGEBAKALLGGSVTVSEKGLTAKVHLGDRAVIA-----	114	
		: : : : : :	:	:
Db	540	GKRIMQRNP-----TEKQSQS--GASVVTFSQSVASTNPLOSSVVKKPKWQRS	587	
		: : : : : :	:	:
Qy	115	-----VSSEQAVRDPLVFRIAGAQVREYTAIILDPOVGYSPKTKSALSDGKTHRK	164	
		: : : : : :	:	:
Db	588	VSREKEPEKDISSNKPLQNRVL-----AEEAEKTNV-----TADNKSQIIQDKKNG	634	
		: : : : : :	:	:
Qy	165	TAPTAESQENONAKALEKTKDDKSANAIV-----KPAVNGKTKHT	203	
		: : : : : :	:	:
Db	635	ATTENSQISESAKKLPTANKSQDTSTEKLSEDILAVASSQITEPSKKPEN-----	689	
		: : : : : :	:	:
Qy	204	VRKETVQTAAAIRPKHLTLEQVADALLKANPNVSAHGRLRAGSVLHIINLRIRKAEP	263	
		: : : : : :	:	:
Db	690	--TAENEKPSQTDIATDELPTOSVEAMSPSPSDAAHQBISEIREIL-----TDKLP	739	
		: : : : : :	:	:

QY	532	-----TPEPESVAQTAEKNKPTVDTDFSDNLPSNNH-----IGTBETASAKP	573
Db	951	AEAGVTEDQDFLGTAKQGVQSPGREPASSIHD--ETLPGGSESEATASDEENRDQP	1008
QY	574	ASPSGLAGFLKA-----SSPETILEKTVAEQTPEELHDFLKUV--YE----	613
Db	1009	EERTATSGYQTQSTIEISSETPMDENSTPRDVMDETNNETESPQEFVNITKEYSLY	1068
QY	614	-----TDVAETAPETPDFWNAADDL-----SAL-----	637
Db	1069	SQEYSKPVWASFNGLSDGSKTDATDGRDYNASASTISPPSMBEDKFSKALDRAVRPEE	1128
QY	638	-----LQPAEAPSVEENITETVAETP-----DENATADDLSALLQP	673
Db	1129	TDVKTGAELDIKQVDERLSPAKSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP	1179
QY	674	SEYPAPVEENAAETVADDLSALLQPAEAPAVEENVTTVAETSDFTHTAADLSALLQPAEV	733
Db	1180	NEIKASAEAGEATAV-----VSPGVTQAVVEE--HCASPEEKTU-----EV	1217
QY	734	PAYVEENVTKVAIEIPDNATADDLSALLQSPQPAVEENAAETILETPDNTSEADALP-	792
Db	1218	VSPQSQVTSAGHTPYQSTPDRKSSH-LPTEV-----TENAQAQVPV	1258
QY	793	--DFLKDGESETVDMWSIYLSEENIPNNAADTSFPSESVGSD--AP-----SEAKYDLAEMYL	844
Db	1259	SPEF-TEAKDENERSSSPMDPEVP-----DSESPLEKVLSPLRSPPLIGSESAY---EDFL	1311
QY	845	EIGDR	849
Db	1312	SADDK	1316

RESULT 12

```

US-10-437-963-150334
; Sequence 150334, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 150334
; LENGTH: 1444
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_50580C.1.pep
US-10-437-963-150334

```

	Query Match	5.9%	Score 260.5;	DB 16;	Length 1444;
	Best Local Similarity	20.5%;	Pred.No. 2e-07;		
	Matches 198;	Conservative 146;	Mismatches 347;	Indels 276;	Gaps 47;
QY	47	SVAAVSFOAHGUGGLNIQSNLDEPFGSGITVTGE---	EAKALLGGSVTVSEKGLTAK	103	
		: :		:	:
Db	330	SLSVKAEETH-----	QSN-----	VATCGCKTPEDDATTREFTVDTKEEQNGS	373
		: :		:	:
QY	104	VHKLGRKAVTAVSSEQAVRPVLVFRIGAGQVREYTAILDPVG-----			147
		: :	: :	:	:
Db	374	VEEMKDAE--AVDTEETVQSSVAF-----	DEAIQDHAAATTNPSSDIQSIIRPEETKGRSDV	427	
		: :	: :	:	:
QY	148	-----YSPKTKSALSDGKTHRTKAPTABSQENQNAKLRKTDKDDSANAAYKPAYNGKTHT	203		

Db	428	KABEVSSQSNVAFDAVQDKVIFSPBPDIRPVKKLEQEBTKEADEA-----SNETNS	481
Qy	204	VRKGETVKQIAAIRPKH/LTEQVADALLKANPNVSAHGLRAGSVLHILNLRIRKAEQP	263
Db	482	AIRSD-LNQEDSIVASKQLETALAEA--AFSNLDOEGGI-ADSESOVADLEEAKELB-	536
Qy	264	KPOTAKPKASTAMPSEPSKQATVEKPVKEPEAKVAPEAKAKPAVRPVPVPAANTAAS	323
Db	537	-----ATETEITHQPPAAVSTELPKEDONSTR-SEPHNDDIQHSLEQDSIEVKDTEAA	588
Qy	324	ETAESAPOGAASAIDTPD-----ETGNVAVSPVEGVSAEEETESGLFGSYLLLAGG	379
Db	589	EIQISOERTIATSKEDAVEDDGTAEQPTCVSOEVQVSEEE-----IEITFAE	426
Qy	380	GAALIALLLRLAQAQRARTBEESVPEEBFDLDDAADG-----VKTEPDNVVEASDVVTV	426
Db	631	-----VKTEPDNVVEASDVVTVVDDGOENNVLTSENIAELQLOGLE	672
Qy	427	VEPAPTEP-----APKD---VNDTL--ALDGESEBEELSAKOTFDVETUTPS	469
Db	673	SEBIKSPETETEAGFHTSHAAPSNDFEVNTACETQDTESEAEI--KETEGTKTESIP	730
Qy	470	NRIDLP-----DSLAAA-----QNGILSGALITQDEETQKRAADWNAISTD--S	513
Db	731	QENISVSESNPDSITAGETTSNQELTIAGSIEASED-----NIDTWTGETIDOSNE	785
Qy	514	VYPETPNPNVPEIV-IDTPPEVSAQTAENKPETVD-----TDFSDN	556
Db	786	VFACEAAGDNIPESVSTADTQSMQBLESEEMKKPELVDSLGTTHQKDDAISQKQOEDN	845
Qy	557	LPS--NNHIGTETASAKAPSPGLAGFLKASPETILEK---TVAEVQTPPELHDFLKV	611
Db	846	PTTCETNEIGSTEVSS-----VSEASEQALAHQSNITQCCEQATEE-----	886
Qy	612	YEYDAVETAPEIPDFNAADDLSALLQPEA--PSV--BENITETVAET-PDFNATADD	666
Db	887	-----SITESPQILITESVD-----MEDTEATEPELVSEQNVSTSEESVPEENAT--	934
Qy	667	LSALLQPEVPA-----VEENAAEIVA--DDLQALQPAEAPAVEENVTVETVAFSDFT	719
Db	935	-----TEPAPHREIQNDGAELTQHDVKAHELPNQSSGA--IVETAEQADLVA	984
Qy	720	A--ADDSAL--LQPAEV-----PAVENTKVVAETPDFNATADDLSAL	760
Db	985	GEPTDDVQEKDLPEEISNTVDGETGASHQTHAAVEDNNTGEV-----ESSVEASEDQAI	1040
Qy	761	LQSEVPFAVENAAEITLETPDFSNTSADALPFLKD--GEERTVDWSIYLSSENIPNN	817
Db	1041	AHQSNITQCCEQATEESLTSEFQILEWESVQD--IKDTEATEGETI-----FQKNIVPT	1093
Qy	818	ADTSFPSESVGSDAPS-----EAKYDLAEMYLE--IGDRDAAAEFTVKLLFEAGDVLKR	870
Db	1094	SEESVPEENATAKEPAPDDREIQNDGAELTKEHGDGVKDEIIPDQSSGAIVEE---TAQE	1149
Qy	871	AQALAE	877
Db	1150	ANLIASE	1156

RESULT 13
US-10-369-493-19525
; Sequence 19525, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B

; CURRENT APPLICATION NUMBER: US/10/369,493
 ; CURRENT FILING DATE: 2003-02-28
 ; PRIOR APPLICATION NUMBER: US 60/360,039
 ; PRIOR FILING DATE: 2002-02-21
 ; NUMBER OF SEQ ID NOS: 47374
 ; SEQ ID NO 19525
 ; LENGTH: 1859
 ; TYPE: PRT
 ; ORGANISM: Myxococcus xanthus
 ; FEATURE:
 ; NAME/KEY: unsure
 ; LOCATION: (1)..(1859)
 ; OTHER INFORMATION: unsure at all Xaa locations
 US-10-369-493-19525

Query Match 5.9%; Score 258; DB 15; Length 1859;
 Best Local Similarity 22.6%; Pred. No. 4.1e-07;
 Matches 209; Conservative 119; Mismatches 366; Indels 230; Gaps 45;

QY	44	IAASVAVAAAPQAHAGLGLNIQSNLDEFF-----SGSITVTGEEAKALLG--GGSV	93
Db	57	VAASAEPTSLASPADLSMEVAPDAPVPTGDALASTEVEPSSVELLDGTADSEP	116
QY	94	TVSEKGLTAKVHKLGDKAVIAVSSEQAVRDPVLFRIGAGAQVREYTAILDPVGSPTK	153
Db	117	VLAHEQSPAPIDDAWASAPLAVETEH-----MDVADAGDIIDVAAMAEPEASQAAD	168
QY	154	SALSDGKTHRTAPTAESEQNQA-KALRKTDKDSANAAYK-----PAYNGKTHVR	205
Db	169	SEIALDSSEMEVSAADSDTTDWDAGATTDTADAWATADHGATDFAASAGDTYBAQ	228
QY	206	KGE-----TVKQIAAIRPKHLTLE--QVADALLKANPNVSAHGRLAGSVLHPLNRIK	259
Db	229	PAETALDTADEASAEAEPTETALDAAHVSDAAVEHASEMAFDA--SAGSE---PGANYSP	283
QY	260	AE--QPKQOTAKPKA--RTASMPSEPSKQATVEKPEAKVAAPAEKAEKPAVRPEVP	315
Db	284	AEIALDASDVAVPEAYPETALTATASDAATAE---AYPAADFAATAEAYPTTEFATAEAY	340
QY	316	PAANTAAGE-----TAESAPOEA--AASAITD--PTDETGNVSEPVSEVSAEETESGL	367
Db	341	PEAFATADVSDASTAEAYPEVALTASEASDTTAAPELALSASDVITSAPEE-----	395
QY	368	FGGSYTLLAGGGAIALILLRLAQSRRARTTESVPE-----BEPDLDDAADGI	420
Db	396	-----LAADASSEVA-----EASPYGSAASAFPEEPAETAFSEGD	432
QY	421	EITAEVETPATPEAPKND--VNDTLALDGESEBELSAKQTFDVTETDP--SNRIDLDF	476
Db	433	ATAAGETYAEAVLDAAEVSDTAVSEHAFATSETSESATAEAYPEVALDADVSDAATMEL	492
QY	477	DSLAAQNGILSGALTQDEETQKRADADWNAIESTDSVVEPET--FNPNVPEIVIDTPE	534
Db	493	DSVETAABATSDAL-----ASSAWG---SIDAAEEFATTEYAADSDVSELAAIETA	540
QY	535	PESVAQTAENKPTVDTFSDNLPNNHIGTETETASAKPSPGLAGFLKASSPTEILEK	594
Db	541	PETAHSTVDHSSEAFD-----GKEVILDA-----TENSEPEIALDT	576
QY	595	T-VAE--VOTPEELHDFLKYVETDAVAETAPETP--DFNAAADDLSALLOPAEAPSVEE	648
Db	577	SDVAESPAQSPSEL-----TLDATDVAESAIEPEAAEISLEASDV-----AEQSAEI	625
QY	649	NI--TETVAETPDFNATADDLSALLOPSEVPAVEENAAEIVADDLS-----	692
Db	626	ALDGESEVEESP-----AQQSVPTLELNAEVSDEPSLDVIDVESAPGHAG	673
QY	693	---ALLQPA-----EAPAVE--ENVVETVAETSDFTAAADDLSALLOPAEAVPENVTKV	744
Db	674	DEASQFAADGGGSEQPTLEVIETGTEATASEVDFPPEA---APFASDVPSLE---IRPV	726
QY	745	AEIPDFNATADDLSALLOPSEVPAVEENAAEITLETDPDNTSEADALP-----DFLKGQ-	798

Db 727 QVGPDLHAETIEVAG--DPOSSARTPSAEEMF-----DLNPNPEAVPLAANBEFLAIGH 780
 QY 799 -----EEETV-----DWSIYLSSEBINPNNADTSPPSESUGSDAPSEAKYDLA 840
 Db 781 PAPTDESVAMAGDAGAGAEILOPEWA-STSQPRAEAAQTWEATS--GEQAAASDAY---- 834
 QY 841 EMYLEIGDRDAAAEVTKLLEAE 864
 Db 835 -----AGWEDPVSETVELQAEWAE 853

RESULT 14
 US-10-369-493-5784
 ; Sequence 5784, Application US/10369493
 ; Publication No. US20030233675A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Cao, Yongwei
 ; APPLICANT: Hinkle, Gregory J.
 ; APPLICANT: Slater, Steven C.
 ; APPLICANT: Goldman, Barry S.
 ; APPLICANT: Chen, Xianfeng
 ; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
 ; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
 ; FILE REFERENCE: 38-10(52052)B
 ; CURRENT APPLICATION NUMBER: US/10/369,493
 ; CURRENT FILING DATE: 2003-02-28
 ; PRIOR APPLICATION NUMBER: US 60/360,039
 ; PRIOR FILING DATE: 2002-02-21
 ; NUMBER OF SEQ ID NOS: 47374
 ; SEQ ID NO 5784
 ; LENGTH: 3507
 ; TYPE: PRT
 ; ORGANISM: Caenorhabditis elegans
 US-10-369-493-5784

Query Match 5.8%; Score 253.5; DB 15; Length 3507;
 Best Local Similarity 18.8%; Pred. No. 1.9e-06;
 Matches 152; Conservative 143; Mismatches 340; Indels 173; Gaps 34;

QY	82	EEAKALLGGSVTVSEKGLTAKVHKLGDKAVIAVSSEQAVRDPVLFRIGAGAQVREYTA	141
Db	1954	EENSSSNSGGEKPTTKGIVS-----STSATSSSETTAEPHVTTISSTTKDMS	2005
QY	142	ILDPVGY-----SPKTSKALSQKTHRTAPTAESEQNQAALRKTDKDSANAAYKPA	196
Db	2006	SKSPENVMTMSSEPEVSTSSKSTTASSETTVSSSTPSSSEAPLTSPPATTEVITESS	2065
QY	197	YNGKTHVRKGETVKQIAAIRPKHLTLEQVADALLKANPNVSAHGRLAGSVLHPLNIN	256
Db	2066	VKS---TTPKEESSEITVKLSKS---PEVTSSVKSSPST-----PSTT	2105
QY	257	RIKAEQPKPTAKPKAETASMPSEPSKQATVEKPEKPEAKVAAPAEKAEKPAVRPEVP	316
Db	2106	SQSVTSVTPETSKSTVLSSEAP-----VTSTSTEVHTSETKPSLSASSTTGTNTST	2159
QY	317	AANTAASATAEAPQEAASAIPTDFTDETGNVSEPVSEVSAEETESGLFGGSYTLILL	376
Db	2160	PSTSSLASVKTSAPGEGTSASVAPVKLSLSLSPDVSQP--STKTFDATESSTVQASSET	2214
QY	377	AGGGAIALILLRLAQSRRARTTESVPEEEDLDDAADGDIETAEVETPATPEPA	436
Db	2215	-SSGTS-----VKGSTPEESHVTKUSITSNP-----SSSVPEVT--SPKSTPTVPE--	2257
QY	437	PKNDVNDTLALDGESEBELSAKQTFDVTETDTPSNR-----IDLDFSLAAQNGILSGALT	492
Db	2258	-----STEQPTST-----TPSQSLTPMNSNEVLTTSEPHVLSLSLS	2295
QY	493	QDETQKRADADWNAIESTDSVVEPETPNNPNVPEIVIDTPEPESVAQTAENKPTVDTD	552
Db	2296	PD---VSGSTTPNNLSSESTVETPKTSS-----EVLNSEPSSTTEAPTLSPDILSTT	2347
QY	553	FSDNLPNNHIGTETETA-----SAKPASPSGL-----AGFLKASSPETILEKTVAEVQTP	602

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 2, 2004, 19:14:41; Search time 46 Seconds
(without alignments)
1840.186 Million cell updates/sec

Title: US-09-743-674-2

Perfect score: 4404

Sequence: 1 MPAGRLPRCPMMTKFTDCT.....EEAEGDVLKRAQAALQELGI 880

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database:

PIR 78:*

1: piri:*

2: pir2:*

3: pir3:*

4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4404	100.0	880	2	G81786
2	4273.5	97.0	875	2	G81209
3	381.5	8.7	919	2	F83257
4	353.5	8.0	1821	2	A82255
5	316	7.2	1822	2	S33441
6	312	7.1	801	2	T29018
7	276.5	6.3	1262	2	T22523
8	272.5	6.2	1829	2	T24583
9	271.5	6.2	1110	2	I51116
10	269	6.1	2364	2	A58577
11	266.5	6.1	1634	2	T26517
12	264	6.0	2055	2	T31110
13	263.5	6.0	2464	1	QRMSPI
14	260.5	5.9	880	2	D89756
15	258.5	5.9	5327	2	T13564
16	257	5.8	1337	2	T30291
17	253.5	5.8	682	2	F83228
18	253.5	5.8	3507	2	T34513
19	253	5.7	1367	1	S48478
20	248	5.6	873	2	A47283
21	248	5.6	971	2	T19431
22	244.5	5.6	734	2	B42680
23	244.5	5.6	2570	2	T17451
24	244	5.5	1229	2	T25697
25	244	5.5	1616	2	G64242
26	243.5	5.5	1558	2	B71603
27	243	5.5	865	2	A47282
28	241	5.5	2187	2	T30826
29	240.5	5.5	1038	2	AG2187

S-layer protein - T18262
I2 protein - Typa
two-component hybr
protein UNC-89 - C
hypothetical prote
hypothetical prote
tCOF1 protein - mo
probable regulator
hypothetical prote
IGA-specific metal
neurofilament prot
neurofilament trip
FmB protein timpo
still frameshift p
hypothetical prote
Bassoon protein - T42730

ALIGNMENTS

RESULT 1

G81786

Neisseria-specific antigen protein, TspA NMA2146 [imported] - Neisseria meningitidis (st

C;Species: Neisseria meningitidis

C;Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 02-Feb-2001

C;Accession: G81786

R;Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Morel
; Holroyd, S.; Jagels, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandream,
Nature 404, 502-506, 2000

A;Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis Z2491.

A;Reference number: A81775; UID:20222556; PMID:10761919

A;Accession: G81786

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-880 <PAR>

A;Cross-references: GB:AL162758; GB:AL157959; NID:g7380672; PIDN:CAB85358.1; PID:g738076

A;Experimental source: serogroup A, strain Z2491

A;Genetics:

A;Gene: tpaA; NMA2146

Query Match 100.0%; Score 4404; DB 2; Length 880;
Best Local Similarity 100.0%; Pred. No. 1.7e-194;
Matches 880; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MPAGRLPRCPMMTKFTDCTRSNRQPPPTHRGYILKNNRQIKLIAASVAASFOAHGL 60

DB 1 MPAGRLPRCPMMTKFTDCTRSNRQPPPTHRGYILKNNRQIKLIAASVAASFOAHGL 60

QY 61 GGLNQSNLDEPFGSGITVTGEEAKALLGGSVTVSEKGLTAKVHKLGDKAVIIVSSEA 120

DB 61 GGLNQSNLDEPFGSGITVTGEEAKALLGGSVTVSEKGLTAKVHKLGDKAVIIVSSEA 120

QY 121 VRDPVLVFRIGAGAVREYTAILDPVGYSPKTSALS DSGKTRKTAPTAESENQAKAL 180

DB 121 VRDPVLVFRIGAGAVREYTAILDPVGYSPKTSALS DSGKTRKTAPTAESENQAKAL 180

QY 181 RKTDKDSANAAVKPAYNGKHTVVRKGETVVKQIAAARPKHILTLEQVADALLKANPNVSA 240

DB 181 RKTDKDSANAAVKPAYNGKHTVVRKGETVVKQIAAARPKHILTLEQVADALLKANPNVSA 240

QY 241 HGLRAGSVLHLPNLRKAEPKQPKTAKPKAETASMPSEPSKQATVEKPKVPEAKVAA 300

DB 241 HGLRAGSVLHLPNLRKAEPKQPKTAKPKAETASMPSEPSKQATVEKPKVPEAKVAA 300

QY 301 PEAKAEKPAVRPEPVPAAATAAETAASAPQEAASAI DTDTGTGNVSPVEQVSAE 360

DB 301 PEAKAEKPAVRPEPVPAAATAAETAASAPQEAASAI DTDTGTGNVSPVEQVSAE 360

QY 361 EETESGLFGGSGYTLLLAGGAALIALLLLRUAQSKRARRTESVPEEPDLDADDDGI 420

DB 361 EETESGLFGGSGYTLLLAGGAALIALLLLRUAQSKRARRTESVPEEPDLDADDDGI 420

QY 421 EITFAEVETPATPEPAPKNDVNDTLALDGESEELS AKQTFDVTDTSPNNRIDLDFDLSA 480
 Db |||||
 QY 421 EITFAEVETPATPEPAPKNDVNDTLALDGESEELS AKQTFDVTDTSPNNRIDLDFDLSA 480
 Db |||||
 QY 481 AAQNGILSGALTQDEETQKADADWNALESTDSVYEPETFPNPNPVEIVDTPEPESVAQ 540
 Db |||||
 QY 481 AAQNGILSGALTQDEETQKADADWNALESTDSVYEPETFPNPNPVEIVDTPEPESVAQ 540
 Db |||||
 QY 541 TAENKPEVTDTDFSDNLPNNHIGTEETASAKPASPGLAGFLKASSPETTILEKTVAEVQ 600
 Db |||||
 QY 541 TAENKPEVTDTDFSDNLPNNHIGTEETASAKPASPGLAGFLKASSPETTILEKTVAEVQ 600
 Db |||||
 QY 601 TPEELHDFLKVYETDAVAETAPETPDFNAAADDLSALLQPAEAPSVSEENITETVAETPDF 660
 Db |||||
 QY 601 TPEELHDFLKVYETDAVAETAPETPDFNAAADDLSALLQPAEAPSVSEENITETVAETPDF 660
 Db |||||
 QY 661 NATADDLSALLQPAEAPSVSEENITETVAETAPETPDFNAAADDLSALLQPAEAPSVSEENITETVAETPDF 720
 Db |||||
 QY 661 NATADDLSALLQPAEAPSVSEENITETVAETAPETPDFNAAADDLSALLQPAEAPSVSEENITETVAETPDF 720
 Db |||||
 QY 721 ADDLSALLQPAEAPSVSEENITETVAETAPETPDFNATADDLSALLQPAEAPSVSEENITETVAETPDF 780
 Db |||||
 QY 721 ADDLSALLQPAEAPSVSEENITETVAETAPETPDFNATADDLSALLQPAEAPSVSEENITETVAETPDF 780
 Db |||||
 QY 781 PDSNTSEADALPDFLKDGEEETVDSIYLSSEENIPNNADTSFPSESVDGSDAPSEAKYDLA 840
 Db |||||
 QY 781 PDSNTSEADALPDFLKDGEEETVDSIYLSSEENIPNNADTSFPSESVDGSDAPSEAKYDLA 840
 Db |||||
 QY 841 EMYLEIGDRDAAETVQKLEAEAGDVLKRAQALAEELGI 880
 Db |||||
 QY 841 EMYLEIGDRDAAETVQKLEAEAGDVLKRAQALAEELGI 880
 Db |||||

RESULT 2

C81209
 tsfA protein NMB0341 [imported] - Neisseria meningitidis (strain MC58 serogroup B)
 C:Species: Neisseria meningitidis
 C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 19-Jan-2001
 C:Accession: C81209
 R:Tettelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, J.A.; Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.A.; r, H.; Qin, H.; Vamathavan, J.; Gill, J.; Scarlato, V.; Maignani, V.; Pizza, M.
 Science 287, 1809-1815, 2000
 A:Authors: Grandi, G.; Sun, L.; Smith, H.O.; Frazer, C.M.; Moxon, E.R.; Rappuoli, R.; V
 A:Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.
 A:Reference number: AB10000; MUID:20175755; PMID:10710307
 A:Accession: C81209
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-875 <TET>
 A:CROSS-references: GB:AE002391; GB:AE002098; NID:g7225561; PIDN:AAF40784.1; PID:g722556
 A:Experimental source: serogroup B, strain MC58
 C:Genetics:
 A:Gene: NMB0341

Query Match 97.0%; Score 4273.5; DB 2; Length 875;
 Best Local Similarity 97.1%; Pred. No. 1.6e-188;
 Matches 858; Conservative 5; Mismatches 8; Indels 13; Gaps 2;

QY 1 MPAGLPRCPNMWTKFTDCTRSNRIOPTTHRGYIILKNNRQIKLIAASVAVASFOAHGL 60
 Db |||||
 QY 1 MPAGLPRCPNMWTKFTDCTRSNRIOPTTHRGYIILKNNRQIKLIAASVAVASFOAHGL 60
 Db |||||
 QY 61 GGLNLTQSNLDEFPSSITVTGEAKALGGGVTVSEKGLTAKVHLGDKAVIAYSSEQA 120
 Db |||||
 QY 61 GGLNLTQSNLDEFPSSITVTGEAKALGGGVTVSEKGLTAKVHLGDKAVIAYSSEQA 120
 Db |||||
 QY 121 VRDPLVFRIGAGAQVREYTAILDPVGYSPKTKSALSDGKTHRKTAFTAESQENAKAL 180
 Db |||||
 QY 121 VRDPLVFRIGAGAQVREYTAILDPVGYSPKTKSALSDGKTHRKTAFTAESQENAKAL 180
 Db |||||
 QY 181 RXTDKDSANAAVKPAYNGKTHTVRKGETVKQIAAIPKHLTLEQVADALLKANPNVSA 240
 Db |||||

Db 181 RXTDKDSANAAVKPAYNGKTHTVRKGETVKQIAAIPKHLTLEQVADALLKANPNVSA 240
 QY |||||
 QY 241 HGRLRAGSVLHILPNLRIKAEQPKPQATKPKAETASMPSEPSKQATVEKPKVEKPAKVAA 300
 Db |||||
 QY 241 HGRLRAGSVLHILPNLRIKAEQPKPQATKPKAETASMPSEPSKQATVEKPKVEKPAKVAA 300
 Db |||||
 QY 301 PEAKAEKPAVRPEVPAAANTAAETAESAPOEAAAASAIPTTDETGNVSEPEVQVSAE 360
 Db |||||
 QY 301 PEAKAEKPAVRPEVPAAANTAAETAESAPOEAAAASAIPTTDETGNVSEPEVQVSAE 360
 Db |||||
 QY 361 BETES----GLFGGSYTLILAGGGAALIALLLRLAQSKARARTEESVPPEEDLDAA 416
 Db |||||
 QY 361 BETESGLFDGLFGGSYTLILAGGGAALIALLLRLAQSKARARTEESVPPEEDLDAA 420
 Db |||||
 QY 417 DDGIEITFAEVETPATPEPAPKNDVNDTLALDGESEELS AKQTFDVTDTSPNNRIDLDF 476
 Db |||||
 QY 421 DDGIEITFAEVETPATPEPAPKNDVNDTLALDGESEELS AKQTFDVTDTSPNNRIDLDF 480
 Db |||||
 QY 477 DSLAAQNGILSGALTQDEETQKADADWNALESTDSVYEPETFPNPNPVEIVDTPEPE 536
 Db |||||
 QY 481 DSLAAQNGILSGALTQDEETQKADADWNALESTDSVYEPETFPNPNPVEIVDTPEPE 540
 Db |||||
 QY 537 SVAOTAEKPKETVDTDFSDNLPNNHIGTEETASAKPASPGLAGFLKASSPETTILEKT 596
 Db |||||
 QY 541 SVAOTAEKPKETVDTDFSDNLPNNHIGTEETASAKPASPGLAGFLKASSPETTILEKT 600
 Db |||||
 QY 597 AEVQTPPELHDFLKVYETDAVAETAPETPDFNAAADDLSALLQPAEAPSVSEENITETVAE 656
 Db |||||
 QY 601 AEVQTPPELHDFLKVYETDAVAETAPETPDFNAAADDLSALLQPAEAPSVSEENITETVAE 660
 Db |||||
 QY 657 TPDFNATADDLSALLQPAEAPSVSEENITETVAETAPETPDFNAAADDLSALLQPAEAPSVSEENITETVAETSD 716
 Db |||||
 QY 661 TPDFNATADDLSALLQPAEAPSVSEENITETVAETAPETPDFNAAADDLSALLQPAEAPSVSEENITETVAETPD 720
 Db |||||
 QY 717 FHTAADDLSALLQPAEAPSVSEENITETVAETAPETPDFNATADDLSALLQPAEAPSVSEENITETVAETAEI 776
 Db |||||
 QY 721 FHTAADDLSALLQPAEAPSVSEENITETVAETAPETPDFNATADDLSALLQPAEAPSVSEENITETVAETAEI 776
 Db |||||
 QY 777 TLETPDSNTSEADALPDFLKDGEEETVDSIYLSSEENIPNNADTSFPSESVDGSDAPSEAK 836
 Db |||||
 QY 772 TLETPDSNTSEADALPDFLKDGEEETVDSIYLSSEENIPNNADTSFPSESVDGSDAPSEAK 831
 Db |||||
 QY 837 YDLAEMLYIEIGDRDAAETVQKLEAEAGDVLKRAQALAEELGI 880
 Db |||||
 QY 832 YDLAEMLYIEIGDRDAAETVQKLEAEAGDVLKRAQALAEELGI 875
 Db |||||

RESULT 3

F83257
 hypothetical protein PA3115 [imported] - Pseudomonas aeruginosa (strain PA01)
 C:Species: Pseudomonas aeruginosa
 C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
 C:Accession: F83257
 R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warriner, P.; Hickey, M.J.; B
 adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lm
 ; Lory, S.; Olson, M.V.
 Nature 406, 959-964, 2000
 A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic path
 A:Reference number: AB2950; MUID:20437337; PMID:10984043
 A:Accession: F83257
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-919 <STO>
 A:CROSS-references: GB:AE004735; GB:AE004091; NID:g9949216; PIDN:AAG06503.1; GSPDB:GNOC
 A:Experimental source: strain PA01
 C:Genetics:
 A:Gene: PA3115

Query Match 8.7%; Score 381.5; DB 2; Length 919;
 Best Local Similarity 23.0%; Pred. No. 3.6e-10;
 Matches 232; Conservative 134; Mismatches 362; Indels 281; Gaps 43;
 QY 41 IKLIAASVAVASFOAHA-GLGGLNIQSNLDFPSSITVTGEAKALGGGVTVS--- 96

Db	8	VRAIAAA-SVLTSGMAHGLGIEITLKSALNQDLDAEIELL---EVRD-IGSGEIVIEPSLAS	63
QY	97	-EKGLTAKVHKU-----GDKAVIAVSSEQAVRDPVLVFRICA-----GAQVR	137
Db	64	PEEFSKAGVDRIYYITDLKFTPVKPKNGKSVIRVTSSKPVQEPYINFLVQVLWPNGRLLR	123
QY	138	EYTAILDPGVGSPKTKSALSDGKTHRKHTAPTAESENQNAKARKTKDKDSANAAYK---	194
Db	124	EYTVLLDDPLVSPQAAASAPAPV---SAPRATG-----APRAPQAPAPVRITTA	169
QY	195	PAYNGKTHVRKETVKQIOAAAIRPKHLTLEQVADALLKANPNVSAHG---RLRAGSVLH	251
Db	170	PAGSDTYRTV-SNDTLWEIAQNRTRDRVSPVQAMLAQFELNFCAFVDGNGINRLKSGQVLR	228
QY	252	IPNLNRIKAEQPKQTKPKAETAS-----MPSEPSK-QAT	286
Db	229	IPTEQOMLERSPREALSQVQAOQNQSWGRNPAAGSAGARQLDATQRNAAGSAPSKVDAT	288
QY	287	-----	286
Db	289	DNLRIVSGEGKASKGADKGGKGDSKAIADTLAVTKESLDSITRENEELQSRMQDLOSQLD	348
QY	287	-VEKPEVKEPKAVA-----APEAKPKPAVRPEVPFAANTPA	321
Db	349	KLQKLIQLKDLQALAKLOGALGEGGAAQPNAALPDASQPNAAAAPAPQPGTPAAAAPT	408
QY	322	ASETAABSAPOEAAASAITDPTDE--TGNVSEPVQVSABETESGLFGGSYTLILLAG-	378
Db	409	APAGEAPAAPQPPVAPPPAPAAEKPPAPNPADAPVQAQEPASFL-----DELLANP	463
QY	379	-----GGAULIALLLRLRLAQSKRAARTERSVP-----EEEPDLDDAADGII-EITF	424
Db	464	LWLAVIGGSALLALLVILMILSRNAQKEKEGAQAFADTGEQEEDALDLGKGDFDGLTL	523
QY	425	AEVETPATPEPAPKNDVNDTLALGSEFEL-----SAKQTFDVEVETTPSNRIDLD	475
Db	524	DEPE-PQVAVAPQVEKTTAQTSDALGSDIYIAYGRFNQAELLQNAIYDEP-QRTDLR	581
QY	476	FDSLAA-AONGILSGALTQDBETQKRADWNATBSTDSVYEPETFPNPNVEIVIDTPE	534
Db	582	LKLMVYAEMGDREGFARQENELREIGCAQ-PQVEQLKSY-----PAMVAVAVA	631
QY	535	PESVAQTAENKPEIVDTDFSDNLPSNNHIGTEETASAKP-ASPSGLAGFLKASSPETILE	593
Db	632	GLAGAKLAQDELDSFSLD-DLSLDDSGH-----AAKPDAAQQLDDAFDLSLDLGLGD	683
QY	594	KTVAEVOPTPE-ELHDFLKVVETDAVAETAPDTPDFNAAADDLSALLQPAEAPSVENITE	652
Db	684	DVQADLKSDSGALDDLTLSDSLDLAASTPADKP-----VDLDFGLDFAE-----	728
QY	653	TVAEPTDFNATADDLSALLQSPSVPAVEENAAEIVADDLSALLQ---PAEAPAVEENVTE	709
Db	729	-LAETPS-QPKHDDLGGFSLDLDAP--EDKLSD---DDFLSLNDEVFAAAPADNEFTLD	781
QY	710	TVAEISDFHTAAD--DLSALLQPAEVPAAVENVKTVAEIPDFNATADDLSALLQSPSV	767
Db	782	TEAAEPEALSIPDDFDLSLADEPTPEPAPEKGEDSFAQLQDEVSAQLDELAS-----	833
QY	768	AVEENAAEITLETSDNT---SEADALPDFDLKGEETVDWMSIYLSEENIPNNADTSPFS	824
Db	834	-----NLDEPKSATPSFSAEDAAVASALDGD-----ADDDDFP	866
QY	825	ESVGSADAPSEAKYDLAEWYLEIGDRDAAETVQKLLFEAGDVLUKRAQ	873
Db	867	LSGADEAAT--KULDAPAYIDMGSEGRADITDEVL--AEGNDSQQAEE	911

C/Accession: A82255
R/Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.;
Hartshorn, D.; Jernolova, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers, J.
1, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
Nature 406, 477-483, 2000
A/Title: DNA Sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*.
A/Reference number: A82035; MUID:20406833; PMID:10952301
A/Accession: A82255
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-1621 <HEI>
A/Cross-references: GB:AE004181; GB:AE003852; MID:g9655454; PIDN:AAF94159.1; GSPDB:GN00:
A/Experimental source: serogroup O1; strain N16961; biotype El Tor
C/Genetics:
A/Gene: VC0998
A/Map position: 1

RESULT 4
A82255

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-1829 <WIL>

A:Cross-references: EMBL:Z49130; PIDN:CAA88964.1; GSPDB:GN00020; CESP:T06D8.1

A:Experimental source: clone T06D8

C:Genetics:

A:Gene: CESP:T06D8.1

A:Map position: 2

A:Introns: 1391/3; 1432/3; 1470/3; 1505/1; 1520/1; 1616/1; 1644/1; 1687/3; 1742/1

Query Match 6.2%; Score 272.5; DB 2; Length 1829;

Best Local Similarity 20.4%; Pred. No. 8.8e-05;

Matches 166; Conservative 121; Mismatches 344; Indels 183; Gaps 27;

QY 65 IQSNLDPEFGSITVTGEEKALLGGGVTVSEKGLTAKVHKLGDKAVIAVSSBOAVRDP 124

DB 726 VESDEBPASSSTSIPTLSK-----DQVTEAS-----GEETTAATASRETT 771

QY 125 VLVPRIAGAGQVREYTAIILDPVGYSPKTKSALSDBGKTHRTAPTAESENQNAKALRTD 184

DB 772 TSAVTEGSGEET-TVAVVSSGEEPPASSST-----SIPTLSKDDQVTEA---SG 818

QY 185 KDSANAAVKPAYNGKTHTVRKGETVVKQIAAIPKPKHULTLEQVADALLKANPNVSAHRL 244

DB 819 EETTTAAATEASEETTTSAVTEGSGEDTTVVAV-----VSSGEG 858

QY 245 RAGSVLHPNLNRIKAEQPKQAKPAETASMPSEPSKQATVEKPV---KPEAKVAAP 301

DB 859 PASSSTSIPT---ELSKDDQVTEASGEETTTAAATEASEETTTSAVTEGSGEETTVAVV 915

QY 302 EAKAEKPAVRPEPAAANTAASATAESAPQEAASAIPTDPTDGTNAVSEFPVQVSAE 361

DB 916 ESSGEEPPASSSTSIPTLSKDDQVTEASGEETTTAAATEASEET---TTSVTEGSGEE 972

QY 362 ETESGLFGGSYTLLAGGGAALIALLLRLAQSKRARTTESVPEEPDLDADDGIE 421

DB 973 TTTSVTEGSG-----GEETTTSAVTEGSGEETTTAAATEASEETTTSAVTEGSGEETTT 1004

QY 422 ITFAVETPATPEPAPKNDVNDTLALDGESEELSAAKOTFDVETDTPSNRDLDFSLAA 481

DB 1005 VTGSEIEIP-----SEESSSTTHD-----PSIPVITPKPSVSS 1039

QY 482 AQNGILSGALTQDEETOKRADWNNAIRSTDSVYEPETFPNPNVPEIVIDTPEPESVAQT 541

DB 1040 TIENVMS-KTSSEAAEKIIIGEHTGKDDAGKEDD---NMPAFVTANPAGTSTTES 1094

QY 542 AENKPEITVDTFDNLPSNNHIGTE-----ETASAKPSPSGLAGFLK 584

DB 1095 AENVTSITGEED--ENIKMAKELGKQFAADLAKIAAKQGVNLTETADAKDSGETAHVEDEQ 1152

QY 585 ASSPETIL--EKTVAEV--QTPEELHDFLKVYETDAVA--ETAPETPDFNAADDLSALL 638

DB 1153 VSSSTESSIGSEETTTVAKETTEEHE--ASGEEDAPAFVTCAPDTSITEASVTSALT 1211

QY 639 QPAEAPSVENITETVAETPDFNATADDLSALLQ---PSEVPVAEENAAEIVADDLSALL 695

DB 1212 DETTVAADDESPTSAGVEQSSAIIIDSATVASEEQTSSSEATSVIESGSE----- 1261

QY 696 QPAEAPVAEEN-VTETVAETSDPHTAADLSALLQPAEVPVAEENVTKTVAEIPDPNATA 754

DB 1262 ---EVTITDENIVTSTVAQ-----LEEGSGITAAESKEDSVT 1296

QY 755 DDLSALLQPSVPVAEENAAETITLTPDSNTSEADALPDFLKGDEETVDWISYLEENI 814

DB 1297 TE--ATSGTTSVSSDGSGETVPNDSETSTSSQSTTDDGGSGVTAES--KDESS 1352

QY 815 PNNADTSPFSSVSGSDAPSEAKYDLAEMYLEIGD 848

DB 1353 TTEAPAFVTSKTSGGDEEDSDPTHEFLTGTIDE 1386

RESULT 9

151116

NF-180 - sea lamprey

C/Species: Petromyzon marinus (sea lamprey)

C/Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 21-Jul-2000

C/Accession: I51116

R/Jacobs, A.J.; Kamholz, J.; Selzer, M.E.

Brain Res. Mol. Brain Res. 29, 43-52, 1995

A/Title: The single lamprey neurofilament subunit (NF-180) lacks multiphosphorylation

A/Reference number: I51116; MUID:95287814; PMID:7770000

A/Accession: I51116

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-1110 <MAC>

A:Cross-references: EMBL:U9361; NID:G632548; PIDN:AAA80106.1; PID:G632549

C:Superfamily: neurofilament triplet H protein

Query Match 6.2%; Score 271.5; DB 2; Length 1110;

Best Local Similarity 21.3%; Pred. No. 5.1e-05;

Matches 181; Conservative 119; Mismatches 330; Indels 219; Gaps 33;

QY 66 QSNLDPEFGSITVTGEEK-----ALLGGGVTVSEKGLT-----AKVHKLGDKAVIA 114

DB 177 QEHLDEEIQRLREKTDDEVLNRNETEALINAFRNKNDVDTSLVRMEMDKRTQSLLDETF 236

QY 115 VSSQAVRDPVLVPRIGAGQVREYTAIILDPVGYSPKTKSALSDBGKTHRTAPTAE 171

DB 237 KKNHEEEVDLELL-----AQIQSTVSVERKDFAVPEITTAALREIRQLGEGQARNIET 289

QY 172 QENQ-NAKALRKTKDSDANAAVKPAYNGKTHTVRK-----GETVKQIAAA 216

DB 290 AEFWPKGFSQTEAAGQNDAIRSAKEEITHEHRKLQMRCTELDALAGTKESLERQ 349

QY 217 IRPKHLT-----LEQVADAL-----LQKPNVSAHGLRAG 247

DB 350 MEERHQSDVGNLQDAQAQLENELNRTKWMARHLREYQDLLNVKMLDIEIAAARV 409

QY 248 SVLHPNLNRIKAEQPKQAKPAETASMPSEPSKQATVEKPV---EKPEAKVAAP 305

DB 410 EEIYR-----SSGFLTPAKP-----PKAPSAKPAKVKVSKKPAEIKVES----- 453

QY 306 EKPAVRPEPVA-ANTAASATAESAPQEAASAIPTDPTDGTNAVSEFPVQVSAEETE 364

DB 454 -----EPISAQDLDLEDLAQEEVWEAKAAPVVAEAEKEEEEEEEEEEEAEAE 506

QY 365 SGLFGGSYTLLAGGGAALIALLLRLAQSKRARTTESVPEEPDLDADDGIEIT 424

DB 507 E-----EEDRGKGEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAE 535

QY 425 AEVETPATPEPAPKNDVNDTLALDGESEELSAAKOTFDVETDTPSNRDLDFSLAAQN 484

DB 536 AEVE-----EAE 585

QY 485 GILSGALTQDEETOKRADWNNAIRSTDSVYEPETFPNPNVPEIVIDTPEPESVAQT 544

DB 586 --AAAKAEVEEEAE 636

QY 545 KPETVDTDFSNLPSNNHIGTEETASAKPSPSGLAGFLKASSPTILEK---TVAEVQT 601

DB 637 EAEEAEAE-----EEVTSK-----KAKTQAEVEEEAEAEAEAEAEAEAEAE 672

QY 602 PEELHDFLKVYETDAVETAPETPDFNAAADDLSALLQPAEAPSV-EENITETVAETPD 660

DB 673 EEEAEAE--EAGEEDVEAESEEEEDSKEAD---AEDEAEAEAEAEAEAEAEAEAEAE 727

QY 661 NATADDLSALLQPSVPVAEENAAEIVADDLSALLQPAEAPVAEENVTKTVAEISD 720

DB 728 EA-----EEAAKSEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAE 773

QY 721 ADDLSALLQPAEVPVAEENVTKTVAE-----PDFNATADDLSALLQPSV 767

DB 774 SDD-----EKP--EEVKESEAPVAEPAKPAKPAKPAKPAKPAKPAKPAKPAK 823

QY 768 AVEENAAETITLTPDSNTSEADALPDFLKGDEETVDWISYLEENIPNNADTSPFSES 827

[illegible]

Db 1287 -----GEARVSVCVTAQVVE-----HCASPEEKL-----EWSPSVCVTSAGHTFYQ 1333

QY 752 ATADDLSALLOFSEVPVAVEENAAEITLETDPDNTSEADALPDFLKDGEEETVDWSIVLSE 811

Db 1334 SPTDEKSSHL-PT-----VSENAQVPV-----SFESEA-----KDENERA-----SLSPMD 1376

QY 812 ENIPNNADTSPSESV-----GSDAPSE 834

Db 1377 BPVP-----DSESPVEKVLSPRLSPPLLGSESPYE 1406

RESULT 14

Db9756

protein T23E7.2b [imprinted] - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 09-Nov-2001

C:Accession: Db9756

R:Anonymous, The C. elegans Sequencing Consortium.

Science 282, 2012-2018, 1998

A:Title: Genome sequence of the nematode C. elegans: a platform for investigation

A:Reference number: A75000; MUID:99069613; PMID:9851916

A:Note: see websites genome.wustl.edu/gsc/C_elegans/ and www.sanger.ac.uk/Projects/C_elegans/

A:Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999

A:Accession: Db9756

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-880 <STO>

A:Cross-references: GB:chr_X; PIDN:AAE71258.1; PID:g2435547; GSPDB:GN00028; CES

C:Genetics:

A:Gene: T23E7.2b

A:Map position: X

Query Match 5.9%; Score 260.5; DB 2; Length 880;

Best Local Similarity 22.0%; Pred. No. 0.00012;

Matches 174; Conservative 89; Mismatches 284; Indels 243; Gaps 36;

QY 172 QENONAKALRKT-----KDSANAAVKPAYNGKTHTVRKGETVKQI-----AAAIRPK 220

Db 31 QDNSNAQLVEEGDQLVTAEADDEKSA-----QVDEDRTEQVPDITGTETVDPM 81

QY 221 HLTLQVADALLKANPNVSAHGLRAGSVLHINLRIKAEQPKQATKPKAETASMPSE 280

Db 82 TTSMDGLD-----LPANDEAQVE-----ESAPADEPAPKEPKQEPAPAEPPA 125

QY 281 PSKQATVEK-PVEKPEAKVAAP-----EAKAKPAVRPEVPVPAANTAAGETAESAPOQAAAS 337

Db 126 PABEAPQEEVPABEPVAAETAPAEAAVEEPVE-----EAPVPEAPEAEETAPAE--EP 179

QY 338 AIDPTDTEGNVSEPVQVSAEETESGLFGSGYTLILLAGGAALIALLLRLLAQSKR 397

Db 180 LIEVAVEEA--PAEPEVAE--EAPPEAAKEEFE-----LQAPPEGKR 217

QY 398 ARR---TEESVPPEEPOLDADDDGEI-----TEAEVETPATPEP 435

Db 218 PRTPNDLNPATDEEPAELEADAGEVPTPTKAASEKQSVPTPKSAAPSEVPPTPKS 277

QY 436 APKNDVNDTIALDGESEELSAKQTPDVE-----DTPSNRIDLDF-----DS 478

Db 278 AAPSEVPPTPKSAAPSEVPLTPKSAQSSVPTPKSAVPSEAPPTPSAAPSVDPTPKS 337

QY 479 LAAQNGILSGALTQDEETQKRADADWNAIET-----DSVYE-----516

Db 338 NGIANGSIPGTPSAVQASAKATPAATPAATATPAATATPAATGTPKADTAFNFEETPAT 397

QY 517 PETFNPNVPEIVDTPESVQAQTAENKPEIVDTPDSDNLPSNNHIGTEETASAKPASP 576

Db 398 PRSSVPATPTESNLITPAPKT-----PKTPTPTPTPTKT-----PRTP 437

QY 577 SGLAGELKASSPETILEKTVAEVQ--TPEELHDFLKVYETDAVAETAPETPDFENAAADDLS 635

Db 438 -----KTPAVVEPEPEVAE-----RPEPVAEPEPE--PEPVAE-----470

QY 636 ALLQPAEAPSVENITETVAETPDFNATADDLSALLQPSSEVPAVEENAAEIVADDLSALL 695

```

471  ---EPAEPAVEE---PAEPE-----PADETAETAE----- 501
696  QPAAEPAVENVTET-VAETSDHHTAADDLSALLOPA-----EVPAAVENTKTVAEIPDF 750
502  --AEPEAVEESIEKTEVVEESAPPAARQSPSPPPARRRQSPSPERQRTSRHAD-RDI 558
751  NATADDLSALLOQSEVPAVENNAEITLETDPDNTSEADALPOFLKDGEB---ETVDWSI 807
559  TSYDEDSYRAVPPRMP-----TATFSWSPPDKQSYTPISP-FVSTANKVRNEYTSRSS 613
808  YLSEENIPNADTSFSPESVGDSPSEAKYDLAEYMLEIGDRDAAAEVQKLEAEARGDV 867
614  YR-----PTNMYTSHFDDIVATGAFSSALY-----STNRLIERSRST 651
868  LKRAQALAE 877
652  RERQAMRSQ 661

RESULT 15
Ti3564
microtubule-associated protein homolog - fruit fly (Drosophila melanogaster)
N:Alternate names: hypothetical protein EG:49E4.1
C:Species: Drosophila melanogaster
C:Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 17-Nov-2000
C:Accession: T13564
R:Spanos, L.; Papagiannakis, G.; Siden-Kiamos, I.; Louis, C.
submitted to the EMBL Data Library, April 1999
A:Description: Sequencing the distal X chromosome of Drosophila melanogaster.
A:Reference number: Z17689
A:Accession: T13564
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-5327 <SPA>
A:Cross-references: EMBL:AL031128; PIDN:CAA20006.1
C:Genetics:
A:Cross-references: FlyBase:FBgn0025392
A:introns: 24/2; 52/3; 104/3; 179/1; 232/1; 1669/3; 2566/1; 4798/3; 5272/1
A>Note: EG:49E4.1
C:Superfamily: Drosophila 576K microtubule-associated protein homolog

Query Match 5.9%; Score 258.5; DB 2; Length 5327;
Best Local Similarity 18.8%; Pred. No. 0.0016;
Matches 173; Conservative 144; Mismatches 325; Indels 277; Gaps 33;

QY 66 QSNLDPEFSGSTVTGEEAKALLGGGVTVSEKGLTAKVHKLGDKAVIAVSQAVRDPV 125
DB 1387 ETSRPESATGSKVDETEQTKS-----KKSPVPSRPESEAKDKK 1424
QY 126 LVFRIGAGQVREYTAILDPCVSPKTKSALSDGKTHRTKPTAESQENONAKALRKTDK 185
DB 1425 SPFASGEASR-----PESVAESVDEAGKASRRRESIAKTHKDESSLDKAKEQSR 1475
QY 186 KDSANAAVKPAYNGKTHTVRKGETVKQIAA---AIRPKHLT-----LEQVADALLKA 234
DB 1476 RESLAESIKP-----ESGIDKESALASKEASRPESVTDKSKEPSRRRESIAESL--- 1523
QY 235 NNVSAGHRLRAGSVIHIPNLNRIKAEQPKQTPAKPKAETASMP-----SE 280
DB 1524 -----KAESTKDEKSAAPPKESARPGSVSVESVKDETEKSKE 1559
QY 281 PSKQATV---EKP-----VEKPEAKP-----AAPEAKAEP-----AV 310
DB 1560 PSRRRESIAESAKPTEFREVSRPESVIDGIDKESAKPESRRDSPSLASKASRPESVLESV 1619
QY 311 RPEPVAANTAASATAE-----SAPQEAASAIPTPTDETG---N 348
DB 1620 KDEPIKSTESRRRESVAESFKADSTKDEKSLTSDIRPESAVENVMAPPKETSRRPES 1679
QY 349 AV-SEPVOVSAEETESLFGGSYTLILLAGGAALIALILLRLAOSKRAARTESVPE 407
DB 1680 AVGSMKDESMSEKPSRRSVKQGA-----AQSRETSR-PASVAE 1717

```

```

QY 408 EEPDLDDAADDGIIETFAEVETP-----ATPEPAKDVNDTLDLDCGESE 453
DB 1718 ---SAXDGADDLKELSRPESTTQKEAGSIKDEKSPLASSEASRPASVAESVKDEAEKSK 1774
QY 454 ELKAKQTFVETTPSNRIOLDPDSLAAQNGILSGALTODEETOKRADADWNAIEGTDS 513
DB 1775 EESRRRESVAESKSLPSK---EASRPASVAESIKDEAKSKESRRRESVAESKSLPS--- 1827
QY 514 VYBPETFPNPNVEIVDTPEP-----ESVAQTA-----ENKPTVDTDFSDNLPN 560
DB 1828 ---KEASRPASVAESIKDEAKSKESRRRESVAESKSLPSKEASRPASVAESIKDEAKS 1884
QY 561 NHIGTEFTASAKPASPSGLAGFLKASSPETILEKTVAEVQTPPELHDFLKYETDAVET 620
DB 1885 KEESRRRESVAESKSLPS---KEASRPASVAESIKDEAKSKESRRRESVAESKSLPS--- 1933
QY 621 AP-ETDPFNAAADDLSALLOPAEAPAVEENITETVAETPDNFNATADDLSALLOQSEVPV 679
DB 1934 SPLPSKEASRPASVAESIKDEAE-KSKEESRRRESVAE-----KSPLPSK 1976
QY 680 EENAAEIVADDLSALLOPAEAPAVEENVTETVAETSDPHTAADDLSALLOQAEVPAVEEN 739
DB 1977 EASRPASVAESIKDEAKSK-----EESRRRESVAES-----PLPSKEAS 2016
QY 740 VTKTVAEIPDNFATADDLSALLOQSEVPDAVEENAAEITLETDPDNTSEADALPDLKGE 799
DB 2017 RPASVAE-----SIKDEAKSKESRRRESVAESK-----PLPSKEASRPASVAESIKDEA 2066
QY 800 EETVDMYILSEENIPNNAUTSFPSSEVGDSPAEAKYDLAEYMLEIGDRDAAAEVQKL 859
DB 2067 EKS-----KEESRRRESVAESKSLPSKEASRPAA-----SVAESVKDEADKSKESRRRESM 2115
QY 860 LEEAEGDVLKRAQALAE 878
DB 2116 AESGKAQSIKGDQSPLEK 2134

```

Search completed: September 2, 2004, 19:21:37
Job time : 61 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 2, 2004, 19:11:46 ; Search time 27 Seconds
(without alignments)
1697.101 Million cell updates/sec

Title: US-09-743-674-2

Perfect score: 4404

Sequence: 1 MPAGRLPRCPMMTKFTDCT.....EEAGDVLKRAQALAEGLI 880

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	276.5	6.3	2468	1	MAPB_HUMAN
2	269	6.1	2459	1	MAPB_RAT
3	263.5	6.0	2464	1	MAPB_MOUSE
4	257	5.8	1337	1	DEXT_STRDO
5	253	5.7	1367	1	AMVH_YEAST
6	244	5.5	1616	1	P200_MYCGE
7	243	5.5	865	1	CPN_DROME
8	240	5.4	704	1	NP14_RAT
9	239	5.4	1664	1	SLPI_CLOTM
10	238	5.4	3664	1	MINT_HUMAN
11	236	5.4	6632	1	UNB9_CABEL
12	232.5	5.3	1849	1	IGA4_HAEIN
13	221.5	5.0	1018	1	HMW1_MYCPN
14	221	5.0	1087	1	NFH_MOUSE
15	220.5	5.0	831	1	NFH_RAT
16	219.5	5.0	8545	1	ANCL_CABEL
17	219	5.0	1781	1	AK12_HUMAN
18	219	5.0	2004	1	MY83_HUMAN
19	218	5.0	1026	1	NFH_HUMAN
20	215.5	4.9	1159	1	NI24_SCHPO
21	215	4.9	1694	1	IGA2_HAEIN
22	214	4.9	1233	1	YF16_YEAST
23	214	4.9	3644	1	MINT_MOUSE
24	213	4.8	1702	1	IGA2_HAEIN
25	210.5	4.8	442	1	BNB_DROME
26	210.5	4.8	1238	1	SBCC_RHOCA
27	210	4.8	3584	1	NBEA_DROME
28	209.5	4.8	3924	1	ANK2_HUMAN
29	209	4.7	1192	1	RTN4_HUMAN
30	207.5	4.7	1411	1	TCOF_HUMAN
31	206	4.7	5147	1	PCLO_HUMAN
32	205	4.7	1861	1	MAP2_RAT
33	202.5	4.6	1385	1	FAT1_SCHPO

34 202 4.6 705 1 ICAL_BOVIN
35 202 4.6 5085 1 PCLO_RAT
36 201.5 4.6 2090 1 N214_HUMAN
37 201.5 4.6 2805 1 MAPA_HUMAN
38 201.5 4.6 3381 1 PGCV_BOVIN
39 201 4.6 778 1 HTR6_HALN1
40 200.5 4.6 1690 1 C190_DROME
41 200 4.5 670 1 VG50_HSV1
42 200 4.5 2774 1 MAPA_RAT
43 197 4.5 1262 1 STNE_DROME
44 197 4.5 1972 1 P531_HUMAN
45 197 4.5 5120 1 PCLO_CHICK

ALIGNMENTS

RESULT 1
MAPB_HUMAN
ID MAPB_HUMAN STANDARD; PRT; 2468 AA.
AC P46821;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Microtubule-associated protein 1B (MAP 1B) [Contains: MAP1 light chain
DE LCL]
DE MAP1B.
GN Homo sapiens (Human).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Fetal brain;
RX MEDLINE=95104835; PubMed=7806212;
RA Lien L.L., Feener C., Fischbach N., Kunkel L.M.;
RT "Cloning of human microtubule-associated protein 1B and the
RT identification of a related gene on chromosome 15.";
RL Genomics 22:273-280(1994).
CC -!- FUNCTION: The function of brain MAPs is essentially unknown.
CC Phosphorylated MAP1B may play a role in the cytoskeletal changes
CC that accompany neurite extension. Possibly MAP1B binds to at least
CC two tubulin subunits in the polymer, and this bridging of subunits
CC might be involved in nucleating microtubule polymerization and in
CC stabilizing microtubules.
CC -!- SUBUNIT: 3 different light chains, LCL1, LCL2 and LCL3, can associate
CC with MAP1A and MAP1B proteins.
CC -!- DOMAIN: Has a highly basic region with many copies of the sequence
CC KKEE and KKEI/V, repeated but not at fixed intervals, which is
CC responsible for the binding of MAP1B to microtubules.
CC -!- PTM: LCL1 is coexpressed with MAP1B. It is a polypeptide generated
CC from MAP1B by proteolytic processing. It is free to associate with
CC both MAP1A and MAP1B. It interacts with the amino-terminal region
CC of MAP1B (By similarity).
CC -!- SIMILARITY: TO MAP1A.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; L06237; AAA18904.1; -
Genew; HGNC:6836; MAP1B.
MIM; 157129; -
GO; GO:0005875; C:Microtubule associated complex; TAS.
InterPro; IPR000102; MAP1B_Neuraxin.
Pfam; PF00414; MAP1B_Neuraxin; 10.
PROSITE; PS00230; MAP1B_NEURAXIN; 6.
Microtubule; Repeat; Phosphorylation.
CHAIN ? 2468 MAP1 LIGHT CHAIN LCL1.
FT

Fri Sep 3 09:14:08 2004

us-09-743-674-2.rsp

FT REPEAT 1878 1894 MAP1B 1.
FT REPEAT 1895 1911 MAP1B 2.
FT REPEAT 1912 1928 MAP1B 3.
FT REPEAT 1929 1945 MAP1B 4.
FT REPEAT 1946 1962 MAP1B 5.
FT REPEAT 1963 1979 MAP1B 6.
FT REPEAT 1997 2013 MAP1B 7.
FT REPEAT 2014 2030 MAP1B 8.
FT REPEAT 2031 2047 MAP1B 9.
FT REPEAT 2048 2064 MAP1B 10.
FT DOMAIN LYS-RICH (HIGHLY BASIC, CONTAINS MANY
SQ SEQUENCE 2468 AA; 270618 MW; 540839CBDF09D461 CRC64;
KKBE AND KKEI/V REPEATS).
Query Match 6.3%; Score 276.5; DB 1; Length 2468;
Best Local Similarity 20.1%; Pred. No. 0.00051;
Matches 185; Conservative 125; Mismatches 333; Indels 277; Gaps 35;
QY 150 PKTSALSDGKT--HRKTAPTAEQENQAKALRTDKDSANAAVKPAYNGKTHTVRK 207
DB 656 PKKEVAKKEDKTPKKEKPKKEVKEVKEVKEVKEVKEVKEVKEVKEVKEVKEV 715
QY 208 ETVKQIAAIR-PKHLTLEQVADALLKANPNVSAHGRILRAGSVLHPIPLNLRIK 266
DB 716 EKEVKEVKEVKEVKEVKEVKEVKEVKEVKEVKEVKEVKEVKEVKEVKEVKEV 753
QY 267 TAKPKAETASMPSEPSKQATVE-----KPVEKPEAKVAAPAEAKAPAVRPE 319
DB 754 ALKPKV-----PKKEEVKQSVAAKPKPKPKPKPKPKPKPKPKPKPKPKPK 804
QY 320 TAASETAAESA-----POEAASAIDTPTDTGNVSEPVQVSAABETESGLFG 373
DB 805 TAAVMAAGIAIGAIPAKLEAERSLSMSPEDLTQFELKABEVDVTDIKP----- 856
QY 374 LLLAGGAALIALLLRLAQKRARRTEESVPEEPDL-----DAADGIEITFAE 429
DB 857 -----QLELIEDEKLKTEPVYAYVIQKEREVTKGPASPDGIGITTEGEG 905
QY 430 PATP---BPAPKNDVNDLAL-----DGE-----450
DB 906 EOTPELEPVEKQGVDDIEKFDEGAGFESSETGDYEKAEKTEAEPEEDEGEHVC 965
QY 451 -----SEBELSAKQTFD-----ETDTPSNRIDLDFSLAAQNGILSGALQDE 497
DB 966 ASKHSPTDEESAKAEADAVIREKRESVAGDDRAEDMDE--AIEKG-----E 1016
QY 498 QKRADADNAISTDSVPEPFNPNYPIVID-----TPEPESVAOT 541
DB 1017 EEEADEEDKAEDAREEYEPKMEADYVMAVVDKAAEAGGAEEOQGVFTLTPKQLG 1076
QY 542 AENKPTVTDPSDNLPSNNH-----IGTEETASAKPASPGLAGFLKA----- 585
DB 1077 PGREPASSIHD--ETLPGGSESEATASDENREDQEEFTATSGYTSQSTIEISSE 1134
QY 586 --SSPETIILEKTVAEVQTPPEELHFLKV--YETDAVAE---TAPETP----- 625
DB 1135 EMSTPRDVMSETNNEETSPSQEFVNIITKYESLSYQRYSKPADVTPLNGFSEG 1194
QY 626 -----DFNAA-----DLSAL-----LOPA 641
DB 1195 TDGKOYNASASTISPPSSMBEEDKFSRALRDVCSVEKASTTLDIKDSISAVSE 1254
QY 642 EAPSVENITETVAETP-----DFNATADDLSALLQPSVEPVAEENAEIVADD- 694
DB 1255 KFSLSGPSPPLEKTPLGERSVNFSLTPNEIKVSAEAVAPSVPEVTQVVEHCAS 1314
QY 695 LQPAEAPAVENVTETVAETSDHFAADDLSALLQPAV-----PAV-----E 740
DB 1315 DKULEVPSQSVTGSAGHTPYQSPDTEKSSHL-PTEVIEKPPAVPVPVSEFSDA 1373
QY 741 TKTV-----BIPDNATADDLSALLQPSVPAV--BENAAEITLETDSNTSEAD 794
DB 1374 RASVSPMDPEVPD---SESPIEKVLSPLRSPPLIGSESAYESFLSADDKASGR 1430

QY 795 LKQGEETVWSIYLSSENPNNADTSPSPESVSGSDAPSEAKYDLARMYLEIGDR----D 850
DB 1431 EKSGKQSPDQVSPVSE-----MTSTSLYQDKQEGKSTDFAPIKEDFGQEKTTDD 1480
QY 851 AAARETVQKLL---EEAEGDV 867
DB 1481 VEAMSSQPALALDERKLDGV 1500
RESULT 2
MAPB RAT STANDARD; PRT: 2459 AA.
ID MAPB RAT STANDARD; PRT: 2459 AA.
AC P15205; Q62958; Q9ER21; Q9QW92;
DT 01-APR-1990 (Rel. 14, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Microtubule-associated protein 1B (MAP 1B) (Neuraxin) [Contains: MAP1
DE light chain LC1].
GN MAP1B.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE OF 1-142 FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=Testis;
RX MEDLINE=96257242; PubMed=8666295;
RA Liu D., Fischer I.;
RT "Isolation and sequencing of the 5' end of the rat microtubule-
RT associated protein (MAP1B)-encoding cDNA.";
RL Gene 172:307-308(1996).
RN [2]
RP SEQUENCE OF 96-2459 FROM N.A., DOMAIN, AND INDUCTION.
RC STRAIN=Sprague-Dawley; TISSUE=Brain, and Glial tumor;
RX MEDLINE=92347374; PubMed=1639092;
RA Zauner W., Kratz J., Staunton J., Feick P., Wiche G.;
RT "Identification of two distinct microtubule binding domains on
RT recombinant rat MAP 1B.";
RL Eur. J. Cell Biol. 57:66-74 (1992).
RN [3]
RP SEQUENCE OF 1541-2459 FROM N.A., AND TISSUE SPECIFICITY.
RC TISSUE=Spinal cord;
RX MEDLINE=90059871; PubMed=2555150;
RA Rientz A., Grenningloh G., Hermans-Borgmeyer I., Kirsch J.,
RA Littauer U.Z., Prior P., Gundelfinger E.D., Schmitt B., Betz H.;
RT "Neuraxin, a novel putative structural protein of the rat central
RT nervous system that is immunologically related to microtubule-
RT associated protein 5.";
RL EMBO J. 8:2879-2888(1989).
RN [4]
RP DEVELOPMENTAL STAGE, AND PHOSPHORYLATION.
RX MEDLINE=97405699; PubMed=9260743;
RA Ma D., Nothias F., Boyne L.J., Fischer I.;
RT "Differential regulation of microtubule-associated protein 1B (MAP1B)
RT in rat CNS and PNS during development.";
RL J. Neurosci. Res. 49:319-332(1997).
CC -I- FUNCTION: The function of brain MAPs is essentially unknown.
CC -I- Phosphorylated MAP1B may play a role in the cytoskeletal changes
CC that accompany neurite extension. Possibly MAP1B binds to at least
CC two tubulin subunits in the polymer, and this bridging of subunits
CC might be involved in nucleating microtubule polymerization and in
CC stabilizing microtubules.
CC -I- SUBUNIT: 3 different light chains, LC1, LC2 and LC3, can associate
CC with MAP1A and MAP1B proteins.
CC -I- TISSUE SPECIFICITY: Nervous system (spinal cord, brain stem,
CC cerebellum and cerebrum). Not expressed in liver, spleen, kidney,
CC heart or muscle.
CC -I- DEVELOPMENTAL STAGE: In cerebral cortex, spinal cord and sciatic
CC nerve levels are high early in development but decrease during
CC postnatal development and are low in adults. In dorsal root
CC ganglia levels remain high throughout development.
CC -I- INDUCTION: By nerve growth factor.

or send an email to license@isb-sib.ch).

[illegible]

```

Db 676 SVNDVSKLETKTTSVNLNHEEIGNFINDVSEKEVQEQPTTQLETDSEFVLPTYQIVE 735
QY 144 D-----PVGYSPKTKSAL-----SDGTHRKRTAPTAEQ-----EN-----174
Db 736 DSFTESAETNEFSEQKOTLEPTISQTEVETSESNVPTVEQTKLFHQDENLFTPLP 795
QY 175 -----QNAKURKTKDSDANAANKVPAKNGKTHTVRKGETVKQIAAIPRKHLEQ 226
Db 796 LDLTETIESNALFDSKPKDESSDSLEQPTF-----KEIKLDS 833
QY 227 VADALLKANGNVSAHGRLAGSVLHIPLNLRKAEQPKQPTA-----KPAETASM 277
Db 834 TVEV-----FOESSQVATEFTVQPEAVFDEIKTQELQPEATTEVDFHQPDPVEQ 889
QY 278 PSFESKQATVEKPKVEKPAKVAAP-----EAKAKPAVRPEVP 316
Db 889 PQBAKFPDVEIPOESSQAFHAFQISDEIKLEKTEAVPHQOLENQSEETVPTTEVT 948
QY 317 AANTAASETAAESAPQEAASAIPTDPTGNAVSEPV-----EVSAAEETESGLFGG 370
Db 949 AFEPETIETQLE-----PSSDQP-----SEPALQNHPEIVTAEVEQ---IFDG 990
QY 371 SYTILLAGGAALIALLLLRQAQSKARTESEVPEEPDLDAAADGI-----EIT 423
Db 991 TKL-----BDLKEANFDNVNNEVQPKETEAEIT 1021
QY 424 F---AEVETPATPEKPNKNDVNDTLALDGESEBELSA-----KOTFDVETDTSNR- 471
Db 1022 FDETKELQETSLEPLSTELKSEATFDNVSEASEAVPEKPKOLETQTEKILEEPKSEP 1081
QY 472 -----IDLDFDSLAAMQNGILSALTQD-----BETQKRADAMNAIESTDSVYEPFNPY 523
Db 1082 VQQLITEASFDTV--RHEAVFDKNQQTQTEGLEBPQVSEAEVVDQTTTDTVGEPEA---1135
QY 524 NPVEIIVDTPEPSVAQT-----AENKPEVTDVDFSDNLPNNHIGTEETASAKPAPSG 579
Db 1136 -----VFDV--QPKETTEVDFDVENQOKVISEPQVQGEAVF--EPSAAKFDSP---1184
QY 580 AGFLKASSPETILE--KTVAEVQ-----TPEELHDFLKVET-----DAVAETAPETPDF 627
Db 1185 VESVQDSQPEFVLEEVQTEIQIQVSEQPEATFTVQPEQTPOEAKFDSVETV-EQPEF 1243
QY 628 N-----AAAD-----LSALLOPAE-----APSVB 647
Db 1244 SSEPTQHVSEASFDEPNYDFDEPNYDFQPSYDGLQPSQYDVBENYDFDEPNYE 1303
QY 648 ENITVETPDNATADDLSALLOPSEVPVAVENAAEIVADDLSALLOPAE-APAVREN 706
Db 1304 ---IESKPEQPEQVEQ-----QGE--AVFEPSEAKFDSFVSVQDSQPEPLLEE- 1352
QY 707 VTETVAETSDFHTAADDLSALLOPAEVP--AVEENVTKTVAEIPDFNATADDLSALLOPS 764
Db 1353 -VQTEIQIQVSEQPEATFTVQPEQTPOEAKFDSVETIQE-PQVSSPFE---VVQPN 1407
QY 765 EYPVAVENAAEITLETDSNTSEADALPDLKDGEEETVDSIYISEEN 813
Db 1408 ----FEERPETVLEEPQADEIQPEA-----SEESLDWELLVGNNS 1445

```

RESULT 7

```

CPN_DROME STANDARD; PRT; 865 AA.
AC Q02910;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Calphorin.
GN CPN OR CAP.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;

```

```

[1]
RN SEQUENCE FROM N.A.
RP STRAIN=Canton-S;
RX MEDLINE=93165729; PubMed=8094559;
RA Martin J.H., Benzer S., Rudnicka M., Miller C.A.;
RT "Calphorin: a Drosophila photoreceptor cell calcium-binding protein.";
RL Proc. Natl. Acad. Sci. U.S.A. 90:1531-1535 (1993).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=Canton-S;
RX MEDLINE=93165730; PubMed=8434015;
RA Ballinger D.G., Xue N., Harshman K.D.;
RT "A Drosophila photoreceptor cell-specific protein, calphorin, binds
calcium and contains a leucine zipper.";
RL Proc. Natl. Acad. Sci. U.S.A. 90:1536-1540 (1993).
CC -I- FUNCTION: Might function as a calcium-sequestering "sponge" to
regulate the amount of free cytoplasmic calcium. It binds 0.3 mole
of Ca(2+) per mole of protein.
CC -I- SUBUNIT: Homodimer (Probable).
CC -I- SUBCELLULAR LOCATION: Cytoplasmic; hypodense compartment.
CC -I- TISSUE SPECIFICITY: Soma and axons of photoreceptor cells of
compound eyes and ocelli.
CC -I- DEVELOPMENTAL STAGE: Expressed early in photoreceptor cell
development.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
DR EMBL; L02111; AAA28405.1; -.
DR EMBL; L05080; AAA28420.1; -.
DR PIR; A47282; A47282.
DR PIR; A47283; A47283.
DR FlyBase; FBgn0010218; Cpn.
DR GO; GO:0005509; F:calcium ion binding; IDA.
KW Calcium-binding.
FT CONFLICT 36 36 A -> AVAPAVVA (IN REF. 2).
FT CONFLICT 43 43 I -> T (IN REF. 2).
FT CONFLICT 64 64 I -> V (IN REF. 2).
FT CONFLICT 76 76 T -> A (IN REF. 2).
FT CONFLICT 100 100 P -> PP (IN REF. 2).
FT CONFLICT 126 127 VQ -> AP (IN REF. 2).
FT CONFLICT 154 154 I -> V (IN REF. 2).
FT CONFLICT 160 160 S -> T (IN REF. 2).
FT CONFLICT 534 534 A -> E (IN REF. 2).
FT CONFLICT 699 699 I -> T (IN REF. 2).
FT CONFLICT 703 703 V -> L (IN REF. 2).
FT CONFLICT 721 721 D -> E (IN REF. 2).
SQ SEQUENCE 865 AA; 84781 MW; 2110417E0B0E7CFE CRC64;

Query Match 5.5%; Score 243; DB 1; Length 865;
Best Local Similarity 24.2%; Pred. No. 0.0034;
Matches 172; Conservative 84; Mismatches 285; Indels 170; Gaps 32;

QY 234 ANPNVSAHGRLAGSVLHIPLNLRKAEQPKQPTAKPAETASMPSEPSKQATVEKPEK 293
Db 60 ATVTTPAPAPIAAASVTPV-----ASVAPVVAAPTTPPAASPVSTPVAVAQIPVAVSA 112
QY 294 PEAKVAPEAKAEKPAVRPEVP--AANTAASETAAESAPQEAASAIPTT-----D 344
Db 113 P---VAPPVAAATPTPVQ---IPVAAPVVATPPVAAASAPTAAVTPVTSVPIASPPVPA 166
QY 345 ETGNAVSEPVQVSEETESGLFGSYTLILAGGAALIALLLLRQAQSKARTEES 404
Db 167 NTVVVAAPVAAVPAAPVVPVAP-----VLAPAVAPAVAPVVAETPAPPVAPVAT 218
QY 405 VPE-EEDPLDD-----AADDGIEITFAEVEPTPATPEKPNKNDVNDTLALDGESEE 454
Db 219 IPECVAPLIPESVVATKPLAAAEPPVVPVATPATFTVVPVAPAAASPHVSVAPAVATVAP 278

```


Db 404 KPAPAAKAVA---TPKQ-----PAGSGQKPSKADSSSE-----BESSSE 443

Qy 648 ENITETVAETDFNATADDLSALLQSPVPAVENAA--EIVADDLSALLQ-----PAEA 700

Db 444 EEATKKSVTTPKARVTAKAAPSL-----PAKQPRAGGDSSESSSEEEKTPPKP 497

Qy 701 PAVEENVETVAETSDFTAAADDLSALLQPAEVPVAVENVTKTVAETPDPFNATADDLSAL 760

Db 498 PAKKKAAGAAVPKPTPVKKAASSESSSSSEDSSEBEK-----KKPKSKATPKPQAG- 550

Qy 761 LQPSVEVPAVENAA---EITLPTDNTSADALP-DFLKQGEETVDMYSIYSEENIP 815

Db 551 -KANGVPASQNGKAGKSESEEEETEONKKAAGTKPGSKRKKNETADEAATPQSKVK 609

Qy 816 NNADTSPSSVSGDAPSEAKYDIAEMYLEIGDRDAAAEVTKLLLEAEGDVLKRA 871

Db 610 LQTNTPFKRKGKGRASSPFRVREEIEVDSR--VADNSFDAKRGAGDWGERA 663

RESULT 9

SLP1_CLOTH STANDARD; PRT: 1664 AA.

AC Q06852;

DT 01-JUN-1994 (Rel. 29, Created)

DT 01-JUN-1994 (Rel. 29, Last sequence update)

DT 01-FEB-1996 (Rel. 33, Last annotation update)

DE Cell surface glycoprotein 1 precursor (Outer layer protein B) (S-layer protein 1).

DE OLBP.

GN Clostridium thermocellum.

OS Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;

OC Clostridium.

OX NCBI_TaxID=1515;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=NCIB 10682;

RX MEDLINE=9320931; PubMed=8458832;

RA Fujino T., Beguin P., Aubert J.-P.;

RT "Organization of a Clostridium thermocellum gene cluster encoding the cellulosomal scaffolding protein CtpA and a protein possibly involved in attachment of the cellulosome to the cell surface.";

RL J. Bacteriol. 175:1891-1899 (1993).

CC -!- SUBUNIT: ASSEMBLED INTO MONO-LAYERED CRYSTALLINE ARRAYS.

CC -!- SUBCELLULAR LOCATION: Cell wall.

CC -!- SIMILARITY: Contains 4 S-layer homology (SLH) domains.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----X67506; CAA47841.1; --

DR PIR: T18262; T18262.

DR InterPro: IPR008965; Cellul_bind.

DR InterPro: IPR001119; SLH.

DR Pfam: PF00395; SLH, 3.

DR PROSITE: PS01072; SLH DOMAIN; 2.

KW Cell wall; S-layer; Signal; Repeat.

FT SIGNAL 1 28

FT CHAIN 29 1664

FT CELL SURFACE GLYCOPROTEIN 1.

FT REPEAT 36 763

FT REPEAT 36 191

FT REPEAT 207 363

FT REPEAT 409 565

FT REPEAT 607 763

FT DOMAIN 771 1377

FT APPROXIMATE TANDEM REPEATS OF

FT T-P-S-D-E-P.

FT GLY/PRO/SER/THR-RICH.

FT DOMAIN 1453 1494

FT SLH 1 (INCOMPLETE).

FT DOMAIN 1495 1565

FT SLH 2.

FT DOMAIN 1566 1625 SLH 3.

FT DOMAIN 1626 1646 SLH 4 (INCOMPLETE).

SQ SEQUENCE 1664 AA; 178194 MW; 5F396695BA9FE74B CRC64;

Query Match 5.4%; Score 239; DB 1; Length 1664;

Best Local Similarity 22.2%; Pred. No. 0.011;

Matches 135; Conservative 78; Mismatches 248; Indels 146; Gaps 28;

Qy 260 AEQPKPTAKPKAETASMPSEPSQATVKEPV-----EKPEAKVAAPAKAKPAVRPEPV 315

Db 785 SDEPTSPDEPTSPDEPTSPDEPTSPDEPTSPDEPTSPDEPTSPDEPTSPDEPTSPDEPT 844

Qy 316 PAANTAASETAASAPQEAASAIIDTDTGNAVSEPV---EQVSAEEETESGLFGSY 372

Db 845 PSEPTSPDEPTSPDEPTSPDEPTSPDEPTSPDEPTSPDEPTSPDEPTSPDEPTSPDEPT 890

Qy 373 TLLAGGGAALIALLLLLRLAQSKRARRTSESVPEEPDLDAAADGDIETFAEVETPA- 431

Db 891 -----EPTSPETPEPIPTDPTSPDEPTSPD--EPTSPDEPTSPD 927

Qy 432 --TP--EPAPKNDVNDTLADGESEELSAKQTFDVTDTTPSNRIDLDFSLAAQNGIL 487

Db 928 EPTSPDEPTSPDEPTSPDEPTSPDEPTSPDEPTSPDEPTSPDEPTSPDEPTSPDEPTSP 969

Qy 488 SGALTQDEETQKRADADWNAIESTDSVYEPETENPVNVEIVDTPEPE---SVAQTAEN 544

Db 970 -----DEPTP--SDEPTSPDEPTSPDEPTSPDEPTSPDEPTSPDEPTSPDEPTSPD 1019

Qy 545 KPETVD--TDFSNLPSNNHIGTEE-TASAKPASPSGLAGFLKASSPETILEKTVAEVQT 601

Db 1020 EPTSPDEPTSPDEPTSPDEPTSPDEPTSPDEPTSPDEPTSPDEPTSPDEPTSPDEPTSP 1077

Qy 602 PEELHFLKVYETDAVETAPETPDFAAADLSALLQPAEAPSVEENITETVAETPDFN 661

Db 1078 PSD-----EPTSPDEPTSPDEPTSPDEPTSPDEPTSPDEPTSPDEPTSPDEPTSPD 1124

Qy 662 ATADDLSALLQSPVPAVENAAEIVADDLSALLQPAEAPAVEENVETVAET-SUPHTA 720

Db 1125 PT-----PSDEPTSPDEPTSPDEPTSPDEPTSPDEPTSPDEPTSPDEPTSPDEPTSP 1170

Qy 721 ADDLSALLQPAEVPVAVENV-TKTVAEIP-----DFNATADDLSALLQSPVPAVE 770

Db 1171 SDEPTSPDEPTSPDEPTSPDEPTSPDEPTSPDEPTSPDEPTSPDEPTSPDEPTSPDEPT 1220

Qy 771 ENAAE-ITLETSDNTSEADALPDFLKQGEETVDMYSIYSEENIPNNADT--SPFSES 827

Db 1221 ETPPEPIPTDPTSPDEPTSPD-----EPTSPDEPTSPDEPTSPDEPTSPDEPTSPDEPT 1270

Qy 828 GSDAPSE 834

Db 1271 PTDTPSD 1277

RESULT 10

MINT HUMAN STANDARD; PRT: 3664 AA.

AC Q96758; Q9H9A8; Q9NH55; Q9UQ01; Q9Y556;

DT 10-OCT-2003 (Rel. 42, Created)

DT 10-OCT-2003 (Rel. 42, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE Mx2-interacting protein (SMART/HDAC1 associated repressor protein).

GN MINT OR SHARP OR KIAA0929.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A., FUNCTION, SUBCELLULAR LOCATION, INDUCTION,

RP RNA-BINDING, AND INTERACTION WITH NCOR2; HDAC1; HDAC2; RBBP4; MBD3;

RP RAR AND MTAIL1.

RC TISSUE=Liver, and Pituitary;

RX MEDLINE=21231190; PubMed=11331609.

RA Shi Y., Downes M., Xie W., Kao H.-Y., Ordentlich P., Tsai C.-C.,

RA Hon M., Evans R.M.;
RT "Sharp, an inducible cofactor that integrates nuclear receptor
RL repression and activation.";
RL Genes Dev. 15:1140-1151(2001).
[2]
RN SEQUENCE FROM N.A.
RA Bird C.;
RL Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.
[3]
RN SEQUENCE OF 294-3664 FROM N.A.
RA Rhodes S., Huckle E.;
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
[4]
RN SEQUENCE OF 793-1595 FROM N.A., AND VARIANT PRO-1091.
RC TISSUE=Embryo, and Teratocarcinoma;
RA Isegaki T., Oca T., Hayashi K., Sugiyama T., Oseki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Takahashi-Fujii A., Hara H.,
RA Tanase T., Nomura Y., Togiya S., Komai F., Hara R., Takeuchi K.,
RA Arita M., Nabekura T., Ishii S., Kawai Y., Saito K., Yamamoto J.,
RA Wakamatsu A., Nakamura Y., Nagahara K., Masuho Y., Oshima A.;
RL "NEDO human cDNA sequencing project";
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
[5]
RN SEQUENCE OF 2002-3664 FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=92246063; PubMed=10231032;
RA Nagase T., Ishikawa K.-I., Suyama M., Kikuno R., Hiroseawa M.,
RA Miyajima T., Tanaka A., Kotani H., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XIII.
RT The complete sequences of 100 new cDNA clones from brain which code
RT for large proteins in vitro.";
RL DNA Res. 6:63-70(1999).
[6]
RN INTERACTION WITH PPAR.
RX MEDLINE=21874127; PubMed=11867749;
RA Shi Y., Hon M., Evans R.M.;
RT "The peroxisome proliferator-activated receptor delta, an integrator
RT of transcriptional repression and nuclear receptor signaling.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:2613-2618(2002).
[7]
RN FUNCTION, AND INTERACTION WITH RBPSUH.
RX MEDLINE=2261914; PubMed=12374742;
RA Oswald F., Kostecka U., Astrahantseff K., Bourtelee S., Dillinger K.,
RA Zechner U., Ludwig L., Wilda M., Hameister H., Knoechel W., Liptay S.,
RA Schmid R.M.;
RT "SHARP is a novel component of the Notch/RBP-Jkappa signalling
RT pathway.";
RL EMBO J. 21:5417-5426(2002).
[8]
RN X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS) OF SPOC DOMAIN.
RX MEDLINE=22777836; PubMed=12897056;
RA Ariyoshi M., Schwabe J.W.R.;
RT "A conserved structural motif reveals the essential transcriptional
RT repression function of Spn proteins and their role in developmental
RT signaling.";
RL Genes Dev. 17:1909-1920(2003).
CC -1- FUNCTION: Essential corepressor protein, which probably regulates
CC different key pathways such as the Notch pathway. Negative
CC regulator of the Notch pathway via its interaction with RBPSUH,
CC which prevents the association between NOTCH1 and RBPSUH, and
CC therefore suppresses the transcriptional activity of Notch
CC signaling. Blocks the differentiation of precursor B cells into
CC marginal zone B cells. Probably represses transcription via the
CC recruitment of large complexes containing histone deacetylase
CC proteins. May bind both to DNA and RNA.
CC -1- SUBUNIT: Interacts with MSX2 (By similarity). Interacts with
CC NCOR2, HDAC2, RBBP4, MBD3 and MTA1L1. Interacts with
CC RBPSUH; this interaction may prevent the interaction between
CC RBPSUH and NOTCH1. Interacts with the nuclear receptors RAR and
CC PPAR. Interacts with RAR in absence of ligand. Bind to the
CC steroid receptor RNA coactivator SRA.
CC -1- SUBCELLULAR LOCATION: Nuclear. Associates with chromatin.
CC -1- TISSUE SPECIFICITY: Expressed at high level in brain, testis,

CC spleen and thymus. Expressed at intermediate level in kidney,
CC liver, mammary gland and skin.
CC -1- INDUCTION: By hormone 17-beta-estradiol (E2).
CC -1- DOMAIN: The RID domain mediates the interaction with nuclear
CC receptors (By similarity).
CC -1- DOMAIN: The SPOC domain, which mediates the interaction with
CC NCOR2, is essential for the repressive activity.
CC -1- SIMILARITY: Belongs to the Spn family.
CC -1- SIMILARITY: Contains 1 RID (receptor interacting) domain.
CC -1- SIMILARITY: Contains 4 RNA recognition motif (RRM) domains.
CC -1- SIMILARITY: Contains 1 SPOC domain.
CC -1- CAUTION: Ref. 2 sequences differ from that shown due to erroneous
CC gene model prediction.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF356524; AAK52750.1; -; ALT SEQ.
CC EMBL; AL034555; CAB85442.1; ALT SEQ.
CC EMBL; AL034555; CAB85444.1; ALT SEQ.
CC EMBL; AL450998; -; NOT ANNOTATED CDS.
CC EMBL; AL096858; CAB51072.1; ALT INIT.
CC EMBL; AK000882; BAA91405.1; ALT INIT.
CC EMBL; AK022949; BAA14324.1; ALT INIT.
CC EMBL; AB023146; BAA76773.1; -;
CC InterPro; IPR000504; RNA_rec_mot.
CC PDB; 1OW1; 19-AUG-03.
CC Pfam; PF00076; rrm; 4.
CC SMART; SM00360; RRM; 4.
CC PROSITE; PS00102; RRM; 4.
CC PROSITE; PS00030; RRM_RNP_1; FALSE_NEG.
CC PROSITE; PS00917; SPOC; 1.
CC Transcription regulation; Repressor; Nuclear protein; DNA-binding;
CC Repeat; RNA-binding; Coiled coil; 3D-structure; Polymorphism.
CC KW
CC FT DOMAIN 1 573
CC FT DOMAIN 6 81
CC FT DOMAIN 335 415
CC FT DOMAIN 438 513
CC FT DOMAIN 517 589
CC FT DOMAIN 688 715
CC FT DOMAIN 977 1004
CC FT DOMAIN 1170 1191
CC FT DOMAIN 1408 1428
CC FT DOMAIN 1496 1529
CC FT DOMAIN 1592 1612
CC FT DOMAIN 1928 1944
CC FT DOMAIN 2201 2707
CC FT DOMAIN 3498 3664
CC FT DOMAIN 2130 2464
CC FT DOMAIN 2709 2870
CC FT DOMAIN 125 277
CC FT DOMAIN 240 325
CC FT DOMAIN 616 810
CC FT DOMAIN 624 697
CC FT DOMAIN 2428 2520
CC FT DOMAIN 3220 3482
CC FT VARIANT 970 970
CC FT VARIANT 1091 1091
CC FT VARIANT 2360 2360
CC FT VARIANT 956 956
CC FT CONFLICT 3664 AA; 402245 MW; 5228C58533ESB27B CRC64;
CC FT SEQUENCE
CC Query Match 5.4%; Score 238; DB 1; Length 3664;
CC Best Local Similarity 20.5%; Pred. No. 0.031;
CC Matches 182; Conservative 119; Mismatches 352; Indels 234; Gaps 38;

```

QY 114 AVSSEQAVRDVLFVFRICAGAVREYTAILD---PVGYSPTKTSALSFGKTRKHTAPTAE 170
DQ 1569 ANSTTDSIQEPVLFH-----SRFELTRMQQKEKEKQKPEVEKQEDTENHPKTPESAP 1624
QY 171 SQENQNAKALRKTDKK-----DSANAUKPAYNGKT-----HTVRKGTVVKQIA 214
DQ 1625 ENKDSKELKTPSVGPPSVVVTVLESAPSALEKTTGDTTEAPLVTEKTVTEPAPVSEAK 1684
QY 215 AAIKPKHLTLQVADALLK--ANPNVSA--HGRLRAGSVLHIPLNLRKAKBAQKPKQTAK 269
DQ 1685 PASEPAPVQLQVLDLPFGADDPDKEAAMPAGVEGS-----SGDQPPYLDK 1734
QY 270 PKAETAS-----MPSEPKQATVKEPKPEAKVAAPKAEPKPAVRPEPVAANTAA 322
DQ 1735 PPTPGASFQSAENVDPEPDSTQPLSKPAQKSE--EANEKPAKPDATADAEDANQKA 1791
QY 323 SETAASAPQEAASADITPT---DETCN-----AVSEPVEQVSAEETE 364
DQ 1792 -EAPESQPPASEDLEVDPPVAADKKPKSKRSKTPVQAAVSVIVEKPVTRKSERIDRE 1850
QY 365 SGLFGGSYTLILLAGGAALIALLLRLAQSRAARTEESVPEEPDLDADAADDGIRITF 424
DQ 1851 -----XKXSNPRGEAQKLLKWEAKKITRTASKNSA 1884
QY 425 AEVETATPEP---APKNDVNDTLADGESEBELSAKQTPDVTETPSNRIDLDLDFSLAA 481
DQ 1885 ADLEHPEPSLPLSRTRRRNRSVYATMGDHNRSVPKPEVE-QPRVTRKRLERELQAAA 1943
QY 482 AQNGILSGALTQDEETQKADWNALESTDSVYEPETNP-----522
DQ 1944 VPTTPRGG---RPPKTRRADEE---EENEAKEPATLTKPPEGRWSPRSQKTAAGGQP 1996
QY 523 --YNPVEIVIDTPEPSV-----AOTAKNKE 547
DQ 1997 GKKGKNEPKVDATRPEATTTEVGPIGVKSSMEPKAAEEAGSEQKRDKADGTRKNPPE 2056
QY 548 TVDTRSDNLPS-----NNHIGTEETASAK----PASPGLAGFLKASSPET 590
DQ 2057 TAPVEVVEKKPAPERKNSKGRSRNSRLAVDKSASLKNVDAAVSPRGA--AQAGER 2112
QY 591 ILEKTVAEVQTPRELHFLKVYETDAVETAPTPFNMAAADLSALLQPAEAPSVE--- 647
DQ 2113 --ESGVAV--SEKSPQKEDGLSGLQSDPVDPEKEDVSGFSPSEATQAKQM 2169
QY 648 --ENITETVAETPDFNATA---DLSALLQPS--EVPVAEENAAETVA-----DLSAL 694
DQ 2170 ELEQAVEHIAKLAESASAYKADAEGLAPEDRDPKPAQASETEALAAIGSIINDISG- 2228
QY 695 LQPAEAPAVEENVETVAETSDFTAAADLSALLQPAE-----VPAVEENVTKTVAE 746
DQ 2229 -EPENFPAPPYPFGES---QTLQPPAG--AQALQPSSEGMETDEAVSGILETEAATES 2282
QY 747 IPDFNATADLSALLQPSVPAVEENAAETITETPSNTSEADALPDFLKD-GEETV--- 803
DQ 2283 RPPVN--ADPPGA--GPTDTKEARNSSETSHVPEAKGSKVEVTLVRKDKGROKTRTS 2338
QY 804 -----DWSIYLSEENIP--NNADTFSPSESVGS-----DAPSEAK 836
DQ 2339 RKRNTNKKVAVPESHVSPESQAQGESPAANEGETTVQHPAPQERK 2385

RESULT 11
UN89 CAEEL
ID UN89 CAEEL STANDARD; PRT; 6632 AA.
AC 001761; Q17362;
DC 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DE 10-OCT-2003 (Rel. 42, Last annotation update)
DE Muscle M-line assembly protein unc-89 (Uncoordinated protein 89).
GN UNC-89 OR C09D1.1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;

```

```

OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A., FUNCTION, AND TISSUE SPECIFICITY.
RC STRAIN=Bristol N2;
RX MEDLINE=96180278; PubMed=8603916;
RA Benian G.M., Tinley T.L., Tang X., Borodovsky M.;
RT "The Caenorhabditis elegans gene unc-89, required for muscle M-line
transduction domains.";
RL J. Cell Biol. 132:835-848 (1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Du Z., Le T.T., Wilson R.;
RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RP REVISIONS.
RA Waterston R.;
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
CC -I- FUNCTION: Structural component of the muscle M-line. Myofilament
lattice assembly begins with positional cues laid down in the
basement membrane and muscle cell membrane. UNC-89 responds to
these signals, localizes, and then participates in assembling an
M-line.
CC -I- TISSUE SPECIFICITY: Localizes to the middle of A-bands.
CC -I- SIMILARITY: Contains 1 DBL-homology (DH) domain.
CC -I- SIMILARITY: Contains 1 fibronectin type III domain.
CC -I- SIMILARITY: Contains 49 immunoglobulin-like C2-type domains.
CC -I- SIMILARITY: Contains 1 PH domain.
CC -I- SIMILARITY: Contains 5 RCD5 domains.
CC -I- SIMILARITY: Contains 1 SH3 domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U33058; AAB00542.1; -
DR EMBL; AF003131; AAB54132.2; -
DR PDB; 1PHO; 20-DEC-00
DR WormPep; C09D1.1; CE30426.
DR InterPro; IPR008957; FN_III-like.
DR InterPro; IPR003961; FN_III.
DR InterPro; IPR007110; IG-like.
DR InterPro; IPR003598; IG_c2.
DR InterPro; IPR003006; IG_MHC.
DR InterPro; IPR001849; PH.
DR InterPro; IPR007850; RCD5.
DR InterPro; IPR000219; RhogEF.
DR InterPro; IPR01452; SH3.
DR Pfam; PF00041; fn3; 1.
DR Pfam; PF00047; ig; 47.
DR Pfam; PF00169; PH; 1.
DR Pfam; PF05177; RCD5; 5.
DR Pfam; PF00621; RhogEF; 1.
DR Pfam; PF00018; SH3; 1.
DR SMART; SM00408; IGC2; 23.
DR SMART; SM00325; RhogEF; 1.
DR SMART; SM00326; SH3; 1.
DR PROSITE; PS50010; DH_2; 1.
DR PROSITE; PS50835; IG LIKE; 49.
DR PROSITE; PS50003; PH DOMAIN; 1.
DR PROSITE; PS50002; SH3; 1.
KW Muscle protein; Immunoglobulin domain; Repeat; SH3 domain;
KW 3D-structure.
FT DOMAIN 63 127 SH3.
FT DOMAIN 152 330 DH.
FT DOMAIN 342 498 PH.
FT DOMAIN 547 633 IG-LIKE C2-TYPE 1.

```


DOMAIN	648	736	IG-LIKE C2-TYPE 2.
DOMAIN	748	838	IG-LIKE C2-TYPE 3.
DOMAIN	946	1033	IG-LIKE C2-TYPE 4.
DOMAIN	1044	1132	IG-LIKE C2-TYPE 5.
DOMAIN	1140	1227	IG-LIKE C2-TYPE 6.
DOMAIN	1272	1315	THR-RICH.
DOMAIN	1375	1475	RCSD 1.
DOMAIN	1479	1585	RCSD 2.
DOMAIN	1597	1695	RCSD 3.
DOMAIN	1700	1799	RCSD 4.
DOMAIN	1800	1860	RCSD 5.
DOMAIN	1982	2067	IG-LIKE C2-TYPE 7.
DOMAIN	2071	2163	IG-LIKE C2-TYPE 8.
DOMAIN	2171	2261	IG-LIKE C2-TYPE 9.
DOMAIN	2269	2359	IG-LIKE C2-TYPE 10.
DOMAIN	2367	2455	IG-LIKE C2-TYPE 11.
DOMAIN	2463	2564	IG-LIKE C2-TYPE 12.
DOMAIN	2563	2651	IG-LIKE C2-TYPE 13.
DOMAIN	2657	2746	IG-LIKE C2-TYPE 14.
DOMAIN	2754	2858	IG-LIKE C2-TYPE 15.
DOMAIN	2887	2980	IG-LIKE C2-TYPE 16.
DOMAIN	2994	3081	IG-LIKE C2-TYPE 17.
DOMAIN	3087	3183	IG-LIKE C2-TYPE 18.
DOMAIN	3189	3280	IG-LIKE C2-TYPE 19.
DOMAIN	3286	3376	IG-LIKE C2-TYPE 20.
DOMAIN	3384	3469	IG-LIKE C2-TYPE 21.
DOMAIN	3482	3572	IG-LIKE C2-TYPE 22.
DOMAIN	3580	3667	IG-LIKE C2-TYPE 23.
DOMAIN	3686	3777	IG-LIKE C2-TYPE 24.
DOMAIN	3817	3908	IG-LIKE C2-TYPE 25.
DOMAIN	3920	4009	IG-LIKE C2-TYPE 26.
DOMAIN	4018	4106	IG-LIKE C2-TYPE 27.
DOMAIN	4109	4201	IG-LIKE C2-TYPE 28.
DOMAIN	4212	4297	IG-LIKE C2-TYPE 29.
DOMAIN	4302	4387	IG-LIKE C2-TYPE 30.
DOMAIN	4400	4485	IG-LIKE C2-TYPE 31.
DOMAIN	4489	4580	IG-LIKE C2-TYPE 32.
DOMAIN	4588	4678	IG-LIKE C2-TYPE 33.
DOMAIN	4681	4771	IG-LIKE C2-TYPE 34.
DOMAIN	4873	4961	IG-LIKE C2-TYPE 35.
DOMAIN	4965	5057	IG-LIKE C2-TYPE 36.
DOMAIN	5067	5160	IG-LIKE C2-TYPE 37.
DOMAIN	5171	5260	IG-LIKE C2-TYPE 38.
DOMAIN	5277	5366	IG-LIKE C2-TYPE 39.
DOMAIN	5383	5472	IG-LIKE C2-TYPE 40.
DOMAIN	5487	5578	IG-LIKE C2-TYPE 41.
DOMAIN	5595	5685	IG-LIKE C2-TYPE 42.
DOMAIN	5701	5790	IG-LIKE C2-TYPE 43.
DOMAIN	5815	5904	IG-LIKE C2-TYPE 44.
DOMAIN	5925	6014	IG-LIKE C2-TYPE 45.
DOMAIN	6038	6130	IG-LIKE C2-TYPE 46.
DOMAIN	6150	6239	IG-LIKE C2-TYPE 47.
DOMAIN	6275	6368	FIBROECTIN TYPE-III.
DOMAIN	6413	6502	IG-LIKE C2-TYPE 48.
DOMAIN	6507	6596	IG-LIKE C2-TYPE 49.
DISULFID	568	621	POTENTIAL.
DISULFID	2908	2975	POTENTIAL.
DISULFID	3015	3065	POTENTIAL.
DISULFID	3707	3759	POTENTIAL.
DISULFID	3826	3890	POTENTIAL.
DISULFID	5092	5157	POTENTIAL.
DISULFID	5298	5350	POTENTIAL.
DISULFID	5508	5560	POTENTIAL.
DISULFID	5616	5669	POTENTIAL.
DISULFID	5722	5764	POTENTIAL.
DISULFID	5836	5901	POTENTIAL.
DISULFID	5946	5998	POTENTIAL.
DISULFID	6036	6171	POTENTIAL.
DISULFID	6421	6486	POTENTIAL.
CONFLICT	2137	2137	A -> P (IN REF. 1).
CONFLICT	2245	2247	AKA -> PKP (IN REF. 1).
CONFLICT	2258	2258	A -> P (IN REF. 1).
CONFLICT	2284	2284	E -> G (IN REF. 1).

FT	CONFLICT	2297	2297	M -> I (IN REF. 1).
FT	CONFLICT	3531	3531	A -> G (IN REF. 1).
FT	CONFLICT	3884	3888	DAGEY -> RRRRI (IN REF. 1).
FT	CONFLICT	3929	3929	A -> V (IN REF. 1).
FT	CONFLICT	5134	5134	A -> P (IN REF. 1).
FT	CONFLICT	5145	5145	T -> S (IN REF. 1).
FT	CONFLICT	5185	5185	G -> A (IN REF. 1).
FT	CONFLICT	5199	5199	K -> N (IN REF. 1).
FT	CONFLICT	5202	5202	L -> F (IN REF. 1).
FT	CONFLICT	5213	5213	F -> L (IN REF. 1).
FT	CONFLICT	6178	6178	A -> G (IN REF. 1).
FT	CONFLICT	6268	6268	K -> E (IN REF. 1).
SQ	SEQUENCE	6632	AA; 731665 MW; 262D3EDD62960E89 CRC64;	

Query Match 5.4%; Score 236; DB 1; Length 6632;
Best Local Similarity 19.1%; Pred. No. 0.076;
Matches 162; Conservative 141; Mismatches 288; Indels 258; Gaps 38;

Qy	40	QIKLIAASVAAVSQAHAGLGLNIOQNLDPFGSGSITVTGEEAKALLGGGVTVSEKG	99
Db	1278	ETSLTSTKTTMTSTTEVTSTVGVVETKES-----SESATTVGGSGGVTB--	1326
Qy	100	LTAQVHKLGDKAVIAVSQAVRDPVLVFRICAGAQVREYTAILDPVGYSPKTSALS--	157
Db	1327	-----GSISVSKIE-----VVSQKDSQTDVREGT-----PKRRVSFAEE	1360
Qy	158	-----DGKTHRKTAFTAESQENQNAKALRTDKKOSANNAVPAVNGKTHTVRKGETV	210
Db	1361	ELPKEVIDSDRKKKSPDKKE-----KSPEKTEE-----KPA-----SPTKKTGEV	1404
Qy	211	KQIAAAIRPKHLTLEQVADALLKANPNVSAHGRLAGSVLHPLNLRKAKQPKQTAKP	270
Db	1405	KS-----PKBKS-----PASPTKKEKS-----PAAEVKSPTKKEKSPSS	1439
Qy	271	KAETASMPSEPSQOATVE-----KPVKEPEAKVAAPAEAKAE-KPAVRPEVPVPAANT	322
Db	1440	PTKKEKSPSTTKTGDEVKESPKSPKKEKSEKEDVKSPVKKEKSPDATNIVEVS	1499
Qy	323	SETAESAPOEAAASAIPTDTETGNVSEPVQVSAEETESGLFGGSYTLILLAGGNA	382
Db	1500	SETTIEKT-ETTTMTTEHSEESRTSVKKEKTEPKVDEKPKS-----	1541
Qy	383	LIALLLLRLAQSKRARRTEESVPEE--EPOLDDAADGIEITFAEVETPATP--	437
Db	1542	-----PTKDKKSPKSEKITEEIKSPVKKEKSPKEVE-----EKPASPTKKEKSP	1584
Qy	438	KNDVNDTLALDGE-----SEELSAKQTFDVETDTPSNRIDLDFDLAAQNGILSGALTQ	493
Db	1585	EKPASPTKSENEVKSPKKEKSPKESVVEELKSPKESPEKADDPKSP-----TK	1636
Qy	494	DEETQKRAADWNAIESTDSVVEPETFPNYPNVEIVIDTPEPESVAQTAENKPEITVDTF	553
Db	1637	KEKSPEKS-----ATEDVKSP-TKKEKSPKV-----EKPPTSPTKKESSPTKKTDD	1682
Qy	554	SDNLPSNNHIGTETASAKPASPSGLAGFLKASSPETILEKTVAEVQTPPEELHDFLKVE	613
Db	1683	EYKSPTKKE-KSPQTVEEKPASPT-----KKEKSPE---KSVVEEVKSPKE	1725
Qy	614	TDVAETAPEPTPDFNAADDLSALLQPAEAPSVVEENITETVAETPDFNATADLSALLOP	673
Db	1726	SPEKAEKPKSPKKEKSPKESA-AEEVKSPKKEKSPKESAE-----EK	1769
Qy	674	SEVPAAVEENAAEIVADDLSALLQPAEAPAVEENVETVAETSDFTHTAADLSALLOPAEV	733
Db	1770	PKSPTKKESSPVKMADD-----EVKSPTKKEKSPKESKVEE-----KPASP	1808
Qy	734	PAVEENVITVAEIPDFNATADLSALLQSEVPAAVEENAAEITLETPTDNTSEADALPD	793
Db	1809	TKKEKTPKESAAE-----ELKSPTK---KEKSPSPTKKTGDEKESKSEKP-	1852
Qy	794	FLKDGEEETVDWISYILSEENIPNNADTSFP-----SESVGSDAPSEAKYDLAEMYLEIGDR	849
Db	1853	-----EEK-----PKSPTPKKSPGSPKPKKSKSPKESPAEKPPAPKL-----TR	1889


```

QY 850 DAAAEVQX 858
Db 1890 DLKLTQVTK 1898

RESULT 12
IGA4_HAEMIN STANDARD; PRT; 1849 AA.
AC P45386;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Immunoglobulin A1 protease precursor (EC 3.4.21.72) (IGA1 protease).
GN IGA.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NHTI HK61;
RX MEDLINE=92234949; PubMed=1373717;
RA Poulsen K., Reinholdt J., Kilian M.;
RT "A comparative genetic study of serologically distinct Haemophilus
RL influenzae type 1 immunoglobulin A1 proteases.";
CC J. Bacteriol. 174:2913-2921(1992).
CC -!- FUNCTION: VIRULENCE FACTOR; CLAVES HOST IMMUNOGLOBULIN A
CC PRODUCING INTACT FC AND FAB FRAGMENTS.
CC -!- CATALYTIC ACTIVITY: Cleavage of immunoglobulin A molecules at
CC certain Pro-Xaa bonds in the hinge region. No small molecule
CC substrates are known.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- DOMAIN: THE SIGNAL PEPTIDE GUIDES THE PRECURSOR TO THE PERIPLASMIC
CC SPACE, AND THE CARBOXY-TERMINAL HELPER DOMAIN ASSOCIATES WITH THE
CC OUTER MEMBRANE TO FORM A PORE FOR EXCRETION OF THE PROTEASE
CC DOMAIN. THE HELPER DOMAIN IS THEN RELEASED BY AUTOPROTEOLYSIS (BY
CC SIMILARITY).
CC -!- SIMILARITY: Belongs to peptidase family S6.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/ or send an email to license@isb-sib.ch).

EMBL; M87491; AAA24968.1; -.
DR MEROPS; S06.001; -.
DR InterPro; IPR006315; Autotransport.
DR InterPro; IPR005546; Autotransporter.
DR InterPro; IPR000710; Peptidase_S6.
DR InterPro; IPR004899; Pertactin.
DR Pfam; PF03797; Autotransporter; 1.
DR Pfam; PF02395; IGA1; 1.
DR Pfam; PF03212; Pertactin; 1.
DR PRINTS; PR00921; IGASERPTASE.
DR TIGRFAMs; TIGR01414; autotrans_barl; 1.
KW Hydrolase; Serine protease; Transmembrane; Zymogen; Signal.
FT SIGNAL 1 25
FT CHAIN 26 1021
FT PROPEP 1022 1849
FT ACT SITE 299 299
SQ SEQUENCE 1849 AA; 202957 MW; 79A7D018C7150AEA CRC64;

Query Match 5.3%; Score 232.5; DB 1; Length 1849;
Best Local Similarity 19.2%; Pred. No. 0.023;
Matches 185; Conservative 140; Mismatches 377; Indels 261; Gaps 41;

QY 35 LKNNRQIKLIARSAVAASPOAHG-----LGGINI-QSNLDEPFSSGSIWTT 80
Db 767 LYSGRNVANITSNITASNAQAQVHI GYKTDGTVCVRSYDTGYVTCHNSLSEKALNSFNPT 826

```

```

QY 81 GERAKA-LLGGGSVTVSEKGLTAKVHKLGDKAVIANSSEQAVRDPVLVFRIGAGAQVREY 139
Db 827 NLRGNVNLTENASFTLIGKANLFGTIQSIGTSQVNLKENS-----WHLTGSNVNQL 878
QY 140 TAILDPVGYSPKTKSALSDGKTHRKTAAPTASQENONAKALKR-----TDKK 186
Db 879 N-----LTNGHHILNAQNDANKVTTNTLVNSLSGNGSYFYWDFTNKK 923
QY 187 DSANAAVKPAYNGKTHV--RKGETVKQIAAAIRPKH--LTLQVADA-----LLKANP 236
Db 924 SNKVVVWKSATGNFTLQVADKTE-----PNHNELTLDASNATRNLEVLTIANG 973
QY 237 NV-----SAHGEL-----PAGSVLHPNLRIKAEQPKQTAKE--PKA 272
Db 974 SVDRGAWKYKLRNVNGRYDLYNPEVKRNTQVDTTITTPNDIQADAPSQSNNEETARV 1033
QY 273 ETASMPSEPSKQATV--EKPEKPEAKVAAPAEKAEKPAVRPVPVPAANTAA----- 322
Db 1034 ETPVPPAPATESAIASEQPETEP-AETAQPAEMEETNTANSTAPKSDTATQTEPNSE 1092
QY 323 ---SET-----AAESAPQEAASAIIDTPTTDTGNAVSEPEVQS 358
Db 1093 SVSEETTEKVAENPPQENETVAKNEQATEPTPQNGEVAKEQPTVEANTQTNEATQSEG 1152
QY 359 ABEETESGLFGGSYTLILAGGGAALLAILLLLLLAQSKARRTEESVPEEPD--LDDAA 416
Db 1153 KTEETQ-----AETKSEPTSVTSVENQPKTVSQST 1185
QY 417 DGIEI---TEARVET-----PATPEAPKNDVNDTLALDGESEELSAKQ 459
Db 1186 EDKVVVEKEKAEKAVETEETQKAPQVTSKEPPKQAEPAPEVPTDTNAEAAQALQCTPTT 1245
QY 460 TDFVETDTPSNRIDLDFSLAAQNG-----ILSGALTQDEETQKRADAMNAIESTDS 513
Db 1246 VAAAEETSPNSKPAEETQOPSEKTAEPVTPVVSSENTATQPTETEETAKVE---KEKTOE 1302
QY 514 VVEPETFPYNPVEIVIDTPEPESVAQTAENKEPTVDTDFSDNLPNSNHIGTEETASAKP 573
Db 1303 V--PQVASQESPKQ---EQPAAKPQAOT---KPAEPA--RENVLTTKNVG-EPQPAQPF 1351
QY 574 ASPS---GLAGFLKASSPETILEKTVAEVQTPPEELHDFLKVYETAETAETPDPFNA 630
Db 1352 QTQSTAVPTTGTAAANSKPAAPQAKQKQTPPEARENSTVNTKEPQSQISATVSTEQPA 1411
QY 631 ADDLSALLQPAEAPSVEENITETVAETDFNATADDLSALLQPSVPAVEENAAEIVADD 690
Db 1412 KETSSNVEQAPENSINTGSATTMTET-----AEKSKDPQME---TVTEND 1454
QY 691 LSALLQPAEAPAVEENVTETVAETSDFTAAADDLSALLQPAEVPVAVEENVTVAEIPDP 750
Db 1455 ----RQP-EANTVADNSVANNSESE--SKSRRRSVSQPKETSABETTASTQ----- 1501
QY 751 NATADDLSALLQPSSE---VPAVEENAAEITLTPDSNTSEADALPDFLKDGEETVDWSEI 807
Db 1502 ETTVDNSVSTPKPSRRRTSRVQVNSYE--PVELPTENAEANAENVQ----- 1545
QY 808 YLSENIIPNNAD-----TSFSESVGSQDAPSEAKYDLAEMYLEIGDRDAAAEVQKLEE 862
Db 1546 --SGNNVANSOPALRNLTSKNTNVLNSAMAKAQF---VALNVG--KAVSQHISGLENN 1597
QY 863 AEG 865
Db 1598 NEG 1600

RESULT 13
HMK1_MYCPN
ID HMK1_MYCPN STANDARD; PRT; 1018 AA.
AC Q50365; Q50348; Q50349; Q9R5R4;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)

```

```
DE protein 1).  
GN HMW1 OR MEN447 OR MP394.  
OS Mycoplasma pneumoniae.  
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.  
NCBI TaxID=2104;  
  
[1] SEQUENCE FROM N.A., AND SEQUENCE OF 1-16.  
RN STRAIN=ATCC 29342 / M129;  
RC MEDLINE=862571877 PubMed=8675025;  
RR Dirksen L.B.; Proft T.; Hilbert H.; Plagens H.; Herrmann R.;  
RA Krause D.C.;  
RT "sequence analysis and characterization of the hmw gene cluster of  
RL Mycoplasma pneumoniae.";  
RN Gene 171:19-25(1996).  
  
[2]  
RP SEQUENCE FROM N.A.  
R RC STRAIN=ATCC 29342 / M129;  
RX MEDLINE=97105985; PubMed=8948633;  
RA Himelreich R.; Hilbert H.; Plagens H.; Firk E.; Li B.-C.;  
RH Herrmann R.;  
RT "Complete sequence analysis of the genome of the bacterium Mycoplasma  
RL pneumoniae";  
RN Nucleic Acids Res. 24:4420-4449(1996).  
  
[3]  
RP SEQUENCE OF 176-182 AND 188-198.  
R RC MEDLINE=92077440; PubMed=1743522;  
RX Krause D.C.; Lee K.K.;  
RA "Juxtaposition of the genes encoding Mycoplasma pneumoniae  
RH cytochrome-accessory proteins HMW1 and HMW3.";  
RT Genes 107:83-89(1991)  
RL -! FUNCTION: COMPONENT OF THE CYTOSKELETON-LIKE STRUCTURE WHICH  
CC STABILIZES THE SHAPE OF THE WALL-LESS MYCOPLASMA. THIS  
CC CYTOSKELETON-LIKE NETWORK OF ACCESSORY PROTEINS CONTAINING HMW  
CC PROTEIN 1 TO 5 ALLOWS THE PROPER ANCHORING OF CYTADHESIN PROTEINS  
CC IN THE MYCOPLASMAL MEMBRANE AT THE ATTACHMENT ORGANELLE (BY  
CC SIMILARITY).  
CC  
CC -! SUBCELLULAR LOCATION: Localizes specifically to the attachment  
CC membrane.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outpost at  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
CC or send an email to license@isb-sib.ch).  
CC -----  
DR EMBL: L38997; AAA61697.1; --  
DR EMBL: AE000038; AAB96042.1; --  
DR EMBL: Z32661; CAA83580.1; --  
DR EMBL: Z32662; CAA83581.1; --  
DR PIR: S73720; S73720.  
DR PhosSite: Q50365; --  
DR Cytaadherence; Structural protein; Complete proteome.  
FT CONFLICT 198 198 E -> I (IN REF. 3).  
SQ SEQUENCE 1018 AA; 112214 MW; 3C3DBE273F9ADB7 CRC64;
```

Query Match 5.0%; Score 221.5; DB 1; Length 1018;
Best local similarity 20.1%; Pred.No. 0.031;
Matches 185; Conservative 128; Mismatches 317; Indels 291; Gaps 45;

DY 112 VIAYSSQAVRDVFLVFRIAGACGVREYTAILEDVPVGSPKTK-----SALS DGKTHRKTPA 167

Db 92 VLAGEQQTEAFDPYLQVNGNEA-----YGYYDAGEWMWGSGFYEGDWISTLP 140

DY 168 TAESOENQ-----NAKALKTDKKDSANAAPVKPYNGKHTTV 204

Db 141 QTAEAREQGFGFDNIETPTTFASEDFGLEADVPAFEVAEPSEVEQEFAAEPIVDVQEVQA 200

DY 205 RK--GE---TVKQLAAAIRPKHLTLAQVADALLK-----ANPNVSASHGLRAGRSLV 250

201 VENVGKTATTATVGGAVGTOTBP-VVVRFIVFSISIQPQPVQEAEMOPEVAVPOLEV---- 255

QY 792 PDPLKDGEEETVDSIYLSEENIPNNADTSFPSESUGSDAFSEAKYDLAEMYLEIGDRDA 851
Db 775 -----REKPKMEAKAKEE-----DKGLPQE-----PSKPKTEKAERSSSTDQKD- 814
QY 852 AAETVQKLEE--AEGD 866
Db 815 -SQPSEKAPEDKAAKGD 830

Search completed: September 2, 2004, 19:18:17
Job time : 42 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 2, 2004, 19:14:12 ; Search time 139 Seconds
(without alignments)
1997.525 Million cell updates/sec

Title: US-09-743-674-2

Perfect score: 4404

Sequence: 1 MPAGRLPRRCPPMNTKFTDCT.....EEAEGDVLKRAQALAQELGI 880

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL 25:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phase:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_rvirus:*
- 16: sp_bacteriaph:*
- 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4404	100.0	880	16 Q9JST3	Q9JST3 neisseria m
2	4273.5	97.0	875	16 Q9K147	Q9K147 neisseria m
3	4270.5	97.0	875	2 O86394	O86394 neisseria m
4	436	9.9	962	16 Q8XX7	Q8XXX7 ralstonia s
5	387.5	8.8	947	16 Q87Y10	Q87Y10 pseudomonas
6	381.5	8.7	919	16 Q9HZA6	Q9HZA6 pseudomonas
7	353.5	8.0	1621	16 Q9KTA5	Q9KTA5 vibrio chol
8	347	7.9	1951	16 Q8DB34	Q8DB34 vibrio vuln
9	340	7.7	911	16 Q8BLE1	Q8BLE1 pseudomonas
10	334.5	7.6	927	2 O87015	O87015 pseudomonas
11	321.5	7.3	717	16 Q886D3	Q886D3 pseudomonas
12	319.5	7.3	2768	5 Q9VC00	Q9VC00 drosophila
13	316	7.2	1822	2 Q07290	Q07290 streptococ
14	315.5	7.2	1786	5 Q9U0P0	Q9U0P0 plasmodium
15	312.5	7.1	673	16 Q8XTC4	Q8XTC4 ralstonia s
16	312	7.1	801	5 Q23635	Q23635 caenorhabdi

17	307	7.0	1380	16 Q87MP0	Q87MP0 vibrio para
18	305	6.9	697	16 Q8PJ23	Q8PJ23 xanthomonas
19	305	6.9	1417	3 Q87LY7	Q87LY7 neurospora
20	295.5	6.7	1110	16 Q8ECR4	Q8ECR4 shewanella
21	295	6.7	764	16 Q82WI4	Q82WI4 nitrosomona
22	294.5	6.7	753	5 Q9U7E7	Q9U7E7 drosophila
23	293.5	6.7	753	5 Q9VLL3	Q9VLL3 drosophila
24	285	6.5	1616	5 Q9VSJ0	Q9VSJ0 drosophila
25	282	6.4	1245	5 Q20684	Q20684 caenorhabdi
26	277	6.3	2262	5 Q9V4P4	Q9V4P4 drosophila
27	276.5	6.3	685	16 Q8P7R3	Q8P7R3 xanthomonas
28	274.5	6.2	1132	5 Q9W475	Q9W475 drosophila
29	272.5	6.2	1763	5 Q8I110	Q8I110 caenorhabdi
30	272.5	6.2	1829	5 Q22248	Q22248 caenorhabdi
31	271.5	6.2	1110	13 Q91255	Q91255 petromyzon
32	269.5	6.1	837	5 Q86BM5	Q86BM5 drosophila
33	267.5	6.1	17352	5 Q95YM2	Q95YM2 procamburus
34	267	6.1	10578	5 Q8ISF5	Q8ISF5 caenorhabdi
35	267	6.1	18519	5 Q8ISF6	Q8ISF6 caenorhabdi
36	267	6.1	18534	5 Q8ISF7	Q8ISF7 caenorhabdi
37	266.5	6.1	1634	5 Q9XW25	Q9XW25 caenorhabdi
38	265.5	6.0	688	2 Q9X4J3	Q9X4J3 ehrlichia c
39	265.5	6.0	773	5 Q9VNX6	Q9VNX6 drosophila
40	264.5	6.0	1238	5 Q7YV27	Q7YV27 trypanosoma
41	264	6.0	638	5 Q24259	Q24259 drosophila
42	264	6.0	2055	2 Q85472	Q85472 abiotrophia
43	261.5	5.9	5412	5 Q9WS96	Q9WS96 drosophila
44	260.5	5.9	880	5 Q17338	Q17338 caenorhabdi
45	259	5.9	1764	2 Q93T34	Q93T34 haemophilus

ALIGNMENTS

RESULT 1

Q9JST3 ID Q9JST3 PRELIMINARY; PRT; 880 AA.

AC Q9JST3; DT 01-OCT-2000 (TrEMBLrel. 15, Created)

DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)

DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)

DE Neisseria-specific antigen protein, TspA.

GN TSPA OR NMA2146.

OS Neisseria meningitidis (serogroup A).

OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;

OC Neisseriaceae; Neisseria.

OX NCBI_TaxID=65699;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=22491 / Serogroup A / Serotype 4A;

RX MEDLINE=20222556; PubMed=10761919;

RA Parkhill J., Achtman M., James K.D., Bentley S.D., Churcher C., Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T., Davies R.M., Davis P., Devlin K., Feltwell T., Hamlin N., Holroyd S., Jagels K., Leather S., Moule S., Mungall K., Quail M.A., Rajandream M.A., Rutherford K.M., Simmonds M., Skelton J., Whitehead S., Spratt B.G., Barrell B.G.;

RT "Complete DNA sequence of a serogroup A strain of Neisseria meningitidis 22491."

RL Nature 404:502-506(2000).

DR EMBL; AL162758; CAB85358.1; --

DR PIR; G81786; G81786.

DR GO; GO:0015998; P:cell wall catabolism; IEA.

DR InterPro; IPR002482; LysM.

DR InterPro; IPR001440; TPR.

DR Pfam; PF01476; LysM; 1.

KW Complete proteome.

SQ SEQUENCE 880 AA; 93153 MW; A1701AFFE849338A CRC64;

Query Match 100.0%; Score 4404; DB 16; Length 880;
Best Local Similarity 100.0%; Pred. No. 5.5e-205;
Matches 880; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MPAGRLPRCPMTKFTDCTRSNRIQPTTHRGYILKNNRQIKLIAASVAVASFOAHAGL 60
 Db 1 MPAGRLPRCPMTKFTDCTRSNRIQPTTHRGYILKNNRQIKLIAASVAVASFOAHAGL 60
 QY 61 GGLNIQSNLDPPFGSGITVTGEEAKALLGGGSGVTSEKGLTAKVHKLGDKAVIASSBQA 120
 Db 61 GGLNIQSNLDPPFGSGITVTGEEAKALLGGGSGVTSEKGLTAKVHKLGDKAVIASSBQA 120
 QY 121 VRDVLVFRIGAGAVREYTAILOPVGYSPKTSKALSOGKTHRTKAPTAESEONAKAL 180
 Db 121 VRDVLVFRIGAGAVREYTAILOPVGYSPKTSKALSOGKTHRTKAPTAESEONAKAL 180
 QY 181 RKTOKOSANAAPKAYNGKTHTVRKGETVQIAAAIRPKHLTLEQVADALLKANPNVSA 240
 Db 181 RKTOKOSANAAPKAYNGKTHTVRKGETVQIAAAIRPKHLTLEQVADALLKANPNVSA 240
 QY 241 HGRIRAGSVLHPIPNLNRIKAEQPKQTAETASMPSEPSKQATVEKPEKPEAKVAA 300
 Db 241 HGRIRAGSVLHPIPNLNRIKAEQPKQTAETASMPSEPSKQATVEKPEKPEAKVAA 300
 QY 301 PEAKAEKPAVRPEVPVPAANTAAETAESAPOEAAASALDPTDGTGNVSEPEVQVSAE 360
 Db 301 PEAKAEKPAVRPEVPVPAANTAAETAESAPOEAAASALDPTDGTGNVSEPEVQVSAE 360
 QY 361 EETESGLFGGSGYTLLAGGGAALIALLLRLAQSKRARRTEESVPEEPDLDAAADGI 420
 Db 361 EETESGLFGGSGYTLLAGGGAALIALLLRLAQSKRARRTEESVPEEPDLDAAADGI 420
 QY 421 EITFAEVTPTATPEPAPKNDVNDTLALDGESEELSAKQTFDVTDTFNSRIDLDFSLA 480
 Db 421 EITFAEVTPTATPEPAPKNDVNDTLALDGESEELSAKQTFDVTDTFNSRIDLDFSLA 480
 QY 481 AAQNGILSGALTQDEETOKRADADWNAIESTDSVVEPETFNPVNPVEIVIDTPEPEVAQ 540
 Db 481 AAQNGILSGALTQDEETOKRADADWNAIESTDSVVEPETFNPVNPVEIVIDTPEPEVAQ 540
 QY 541 TAENKPEVTVDITFNSNHNHIGTEETASAKPSPGLAGFLKASSPETILEKTVAEQ 600
 Db 541 TAENKPEVTVDITFNSNHNHIGTEETASAKPSPGLAGFLKASSPETILEKTVAEQ 600
 QY 601 TPEELHDFLKVYETDAVETAPETDFNAAADLALDGESEELSAKQTFDVTDTFNSRIDLDFSLA 660
 Db 601 TPEELHDFLKVYETDAVETAPETDFNAAADLALDGESEELSAKQTFDVTDTFNSRIDLDFSLA 660
 QY 661 NATADDLSALLQPSSEVPAVEENAAEIVADDLSALLQPAEAPAEVNTVETASDFHTA 720
 Db 661 NATADDLSALLQPSSEVPAVEENAAEIVADDLSALLQPAEAPAEVNTVETASDFHTA 720
 QY 721 ADDLSALLQPAEVPVPAVEENVTKTVAEIPDFNATADDLSALLQPSSEVPAVEENAAEITLET 780
 Db 721 ADDLSALLQPAEVPVPAVEENVTKTVAEIPDFNATADDLSALLQPSSEVPAVEENAAEITLET 780
 QY 781 PDSNTSEADALPDFLKDGEEETVDSIYI SEENIPNNADTSPPSESVGSDAPSEAKYDLA 840
 Db 781 PDSNTSEADALPDFLKDGEEETVDSIYI SEENIPNNADTSPPSESVGSDAPSEAKYDLA 840
 QY 841 EMYLEIGDRDAAAETVQKLLLEAEGDVLKRAQALAEELGI 880
 Db 841 EMYLEIGDRDAAAETVQKLLLEAEGDVLKRAQALAEELGI 880

RESULT 2

Q9K147
 ID Q9K147 PRELIMINARY; PRT; 875 AA.
 AC Q9K147;
 DT 01-OCT-2000 (T-EMBLrel. 15, Created)
 DT 01-OCT-2000 (T-EMBLrel. 15, Last sequence update)
 DT 01-JUN-2003 (T-EMBLrel. 24, Last annotation update)
 DE TspA protein.
 GN NMB0341.
 OS Neisseria meningitidis (serogroup B).
 OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
 OC Neisseriaceae; Neisseria.

NCBI_TaxID=491;
 [1]
 RN SEQUENCE FROM N.A.
 RC STRAIN=MC58 / Serogroup B;
 RA MEDLINE=20175755; PubMed=10710307;
 RA Tettelin H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,
 RA Eisen J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.J.,
 RA Nelson W.C., Gwinn M.L., DeBoy R., Peterson J.D., Hickey E.K.,
 RA Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,
 RA Mason T., Ciecko A., Parksey D.S., Blair E., Cittone H., Clark E.B.,
 RA Cotton M.D., Uterback T.R., Khouri H., Qin H., Vamathevan J.,
 RA Gill J., Scarlato V., Massignani V., Pizzo M., Grandi G., Sun L.,
 RA Smith H.O., Fraser C.M., Moxon E.R., Rappuoli R., Venter J.C.;
 RT "Complete genome sequence of *Neisseria meningitidis* serogroup B strain
 MC58";
 RL Science 287:1809-1815(2000).
 DR EMBL; AE002391; AAF40784.1; --
 DR PIR; C81209; C81209.
 DR TIGR; NMB0341; --
 DR GO; GO:0016998; P:cell wall catabolism; IEA.
 DR InterPro; IPR002482; LysM.
 DR InterPro; IPR001440; TPR.
 DR Pfam; PF01476; LysM; 1.
 KW Complete proteome.
 SQ SEQUENCE 875 AA; 92488 MW; 1F921520C167D090 CRC64;

Query Match 97.0%; Score 4273.5; DB 16; Length 875;
 Best Local Similarity 97.1%; Pred. No. 1.1e-198;
 Matches 858; Conservative 5; Mismatches 8; Indels 13; Gaps 2;

QY 1 MPAGRLPRCPMTKFTDCTRSNRIQPTTHRGYILKNNRQIKLIAASVAVASFOAHAGL 60
 Db 1 MPAGRLPRCPMTKFTDCTRSNRIQPTTHRGYILKNNRQIKLIAASVAVASFOAHAGL 60
 QY 61 GGLNIQSNLDPPFGSGITVTGEEAKALLGGGSGVTSEKGLTAKVHKLGDKAVIASSBQA 120
 Db 61 GGLNIQSNLDPPFGSGITVTGEEAKALLGGGSGVTSEKGLTAKVHKLGDKAVIASSBQA 120
 QY 121 VRDVLVFRIGAGAVREYTAILOPVGYSPKTSKALSOGKTHRTKAPTAESEONAKAL 180
 Db 121 VRDVLVFRIGAGAVREYTAILOPVGYSPKTSKALSOGKTHRTKAPTAESEONAKAL 180
 QY 181 RKTOKOSANAAPKAYNGKTHTVRKGETVQIAAAIRPKHLTLEQVADALLKANPNVSA 240
 Db 181 RKTOKOSANAAPKAYNGKTHTVRKGETVQIAAAIRPKHLTLEQVADALLKANPNVSA 240
 QY 241 HGRIRAGSVLHPIPNLNRIKAEQPKQTAETASMPSEPSKQATVEKPEKPEAKVAA 300
 Db 241 HGRIRAGSVLHPIPNLNRIKAEQPKQTAETASMPSEPSKQATVEKPEKPEAKVAA 300
 QY 301 PEAKAEKPAVRPEVPVPAANTAAETAESAPOEAAASALDPTDGTGNVSEPEVQVSAE 360
 Db 301 PEAKAEKPAVRPEVPVPAANTAAETAESAPOEAAASALDPTDGTGNVSEPEVQVSAE 360
 QY 361 EETESGLFGGSGYTLLAGGGAALIALLLRLAQSKRARRTEESVPEEPDLDAAADGI 420
 Db 361 EETESGLFGGSGYTLLAGGGAALIALLLRLAQSKRARRTEESVPEEPDLDAAADGI 420
 QY 421 EITFAEVTPTATPEPAPKNDVNDTLALDGESEELSAKQTFDVTDTFNSRIDLDFSLA 480
 Db 421 EITFAEVTPTATPEPAPKNDVNDTLALDGESEELSAKQTFDVTDTFNSRIDLDFSLA 480
 QY 477 DSLAAQNGILSGALTQDEETOKRADADWNAIESTDSVVEPETFNPVNPVEIVIDTPEPE 536
 Db 477 DSLAAQNGILSGALTQDEETOKRADADWNAIESTDSVVEPETFNPVNPVEIVIDTPEPE 536
 QY 537 SVAQTAENKPEVTVDITFNSNHNHIGTEETASAKPSPGLAGFLKASSPETILEKTVAE 600
 Db 537 SVAQTAENKPEVTVDITFNSNHNHIGTEETASAKPSPGLAGFLKASSPETILEKTVAE 600
 QY 597 AEVQTPPEELHDFLKVYETDAVETAPETDFNAAADLALDGESEELSAKQTFDVTDTFNSRIDLDFSLA 656
 Db 597 AEVQTPPEELHDFLKVYETDAVETAPETDFNAAADLALDGESEELSAKQTFDVTDTFNSRIDLDFSLA 656
 QY 601 AEVQTPPEELHDFLKVYETDAVETAPETDFNAAADLALDGESEELSAKQTFDVTDTFNSRIDLDFSLA 660
 Db 601 AEVQTPPEELHDFLKVYETDAVETAPETDFNAAADLALDGESEELSAKQTFDVTDTFNSRIDLDFSLA 660


```

QY 657 TPDNATADLSALLQSPSEVPVAVENAAEIVADDLSALLQPAEAPAVEENVETETVAETSD 716
Db 661 TPDFNATADLSALLQSPSEVPVAVENAAEIVADDLSALLQPAEAPAVEENVETETVAETPD 720
QY 717 FHTAADLSALLQPAEAPAVEENVTKVAEIPDFNATADLSALLQSPSEVPVAVENAAEI 776
Db 721 FNATADLSALLQSEAPAVEENAAETV-----ADDLSALLQPAEAPAVEENAAEI 771
QY 777 TLETPDSNTSEADALPDFLKDGEETVDMWSIYLSEENIPNNADTSFSESVDGSDAPSEAK 836
Db 772 TLETPDSNTSEADALPDFLKDGEETVDMWSIYLSEENIPNNADTSFSESVDGSDAPSEAK 831
QY 837 YDLAEWYLEIGDRDAAEIVQKLEEAEGDVLKRAQALAEELGI 880
Db 832 YDLAEWYLEIGDRDAAEIVQKLEEAEGDVLKRAQALAEELGI 875

RESULT 3
O86394
ID O86394 PRELIMINARY; PRT; 875 AA.
AC O86394;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE TspA protein.
GN TSPA.
OS Neisseria meningitidis.
OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
OC Neisseriaceae; Neisseria.
OX NCBI_TaxID=487;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=B:15;P1.16; PubMed=10377136;
RA Kizil G., Todd I., Atta M., Borriello S.P., Ait-Tahar K.,
RA Ala'Aldien D.A.A.;
RT "Identification and characterisation of TspA, a major CD4+ T-cell and
RT B-cell stimulating Neisseria-specific antigen.";
RL Infect. Immun. 67:3533-3541(1999).
DR EMBL; AJ010113; CAA09002.2; -
DR GO; GO:0016998; P:cell wall catabolism; IEA.
DR InterPro; IPR002482; LysM.
DR Pfam; PF01476; LysM; 1.
DR InterPro; IPR001440; TPR.
SQ SEQUENCE 875 AA; 92548 MW; 1F920E217A677091 CRC64;

Query Match
Best Local Similarity 97.0%; Score 4270.5; DB 2; Length 875;
Matches 858; Conservative 4; Mismatches 9; Indels 13; Gaps 2;

QY 1 MPAGRLPRRCPMWTKFTDCTRSNRIPQPHRGVYILKNNRQIKLIAASVAVASFOAHAGL 60
Db 1 MPAGRLPRRCPMWTKFTDCTRSNRIPQPHRGVYILKNNRQIKLIAASVAVASFOAHAGL 60
QY 61 GGLINTQSNLDEPFSGSITVTGEAKALGGSVTVSEKGLTAKVHKLGDKAVIAYSSEA 120
Db 61 GGLINTQSNLDEPFSGSITVTGEAKALGGSVTVSEKGLTAKVHKLGDKAVIAYSSEA 120
QY 121 VRDPLVFRIGAGAVREYATLDPVGYSPKTSALSDGKTHRKAPTAESEONAKAL 180
Db 121 VRDPLVFRIGAGAVREYATLDPVGYSPKTSALSDGKTHRKAPTAESEONAKAL 180
QY 181 RKTDKKDSANAAVKPAYNGKTHVTKGETVKQIAAAIRPKHUTLBOVDALLKANPNVSA 240
Db 181 RKTDKKDSANAAVKPAYNGKTHVTKGETVKQIAAAIRPKHUTLBOVDALLKANPNVSA 240
QY 241 HGRLRAGSVLHLPNLRKAEQPKQTAQKPAETASMPSEFSKQATVEKPKPEAKVAA 300
Db 241 HGRLRAGSVLHLPNLRKAEQPKQTAQKPAETASMPSEFSKQATVEKPKPEAKVAA 300
QY 301 PEAKAEKPAVRPEPVPAAANTAASETAESAQAASAIPTDPTGNAVSEPVQVSAE 360

```

RESULT 4

```

Q8XXX7
ID Q8XXX7 PRELIMINARY; PRT; 962 AA.
AC Q8XXX7;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Probable transmembrane protein.
GN RSC1986 OR RS03415.
OS Ralstonia solanacearum (Pseudomonas solanacearum).
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Burkholderiaceae; Ralstonia.
OX NCBI_TaxID=305;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=CM11000;
RX MEDLINE=21681879; PubMed=11823852;
RA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,
RA Ariat M., Billaut A., Brottier P., Camus J.C., Cattolico L.,
RA Chandler M., Choisme N., Claudel-Renard C., Cunnac S., Demange N.,
RA Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schiex T.,
RA Sigquier P., Thebault P., Whalen M., Wincker P., Levy M.,
RA Weissenbach J., Boucher C.A.;
RT "Genome sequence of the plant pathogen Ralstonia solanacearum.";
RL Nature 415:497-502(2002).
DR EMBL; AL646067; C:intracellular; IEA.
DR GO; GO:0005622; C:intracellular; IEA.
DR GO; GO:0005840; C:ribosome; IEA.
DR GO; GO:0003735; F:structural constituent of ribosome; IEA.
DR GO; GO:0016998; P:cell wall catabolism; IEA.
DR GO; GO:0006412; P:protein biosynthesis; IEA.
DR InterPro; IPR002482; LysM.
DR InterPro; IPR001854; Ribosomal_L29.
DR InterPro; IPR001440; TPR.

```

DR InterPro: IPR008941; TPR-like.
 DR Pfam: PF01476; IYEM: 1.
 DR PROSITE: PS00579; RIBOSOMAL_L29; 1.
 KW Complete proteome.
 SQ SEQUENCE 962 AA; 99116 MW; EDPF380852P0730 CRC64;
 Query Match 9.9%; Score 436; DB 16; Length 962;
 Best Local Similarity 22.8%; Pred. No. 5.2e-13;
 Matches 238; Conservative 147; Mismatches 351; Indels 306; Gaps 41;
 QY 46 ASVAVAASF-----QAH-AGLGELNQSLNDEPFSGSITVIG---EEAKALLG--GGSV 93
 DB 19 SAVAIAAASLLLIQPAHAAGFGALHVRSSLGQPLQAEIDLGSVTEEEAQNVLAKLASPD 78
 QY 94 TVSEKGLT-----AKVHKLGDKAVIAVSSEQVRDP---VLVFRIGAGQV--REY 139
 DB 79 AYQAGUTYNPVTSLRSLVRQSGSYVVRVISAQPVAEFPVILVDLITWASGRVSEAY 138
 QY 140 TAILDPVGYSPKTKS-----ALSDGKTHRKTAFTAESQE-----NONAKALRKTDKK 186
 DB 139 TFLDPAGSNTPNFAPTPPVQATTGCVVDSTPAPVAAAPQAPAAAPRAAPARQAAR 198
 QY 187 DSANAAVPAVNGKTHVTKGETVKQIAA-AIRPKH-LTLEQVADALLKANPNVSAHG-- 242
 DB 199 PQADAAAAPSPSGAGYTVQRGDSLYDVASNAVQGDGVSLDQMLLALYRNNPKAFIGNI 258
 QY 243 -RLRAGSVLHLPN---LNRIKAEQPKQ-----TAKPKA 272
 DB 259 NRLTGTSLVLTVPFAAEAKQVSSREARREVVAGTSGFAGYRSLRATAAEANAATDTSARQ 318
 QY 273 ETASMPSEPSKQAT-----VEKPVKEPKAVA 299
 DB 319 QSGSVARVQVQATPSASERDELRLSKADRTGKAAATAGARAEELVAKHALKEMESRVA 378
 QY 300 APE-----AKAEKAVRPEVP-----AANTAASETAESAPO 332
 DB 379 QLEKNLSEMOHLIEVKNAELAKAAAKPAGAAPSPATAAPAVTAANAPAPSASATPA 438
 QY 333 E---NAASAIPT-----PTDETGNVSEVPVQVSA----- 359
 DB 439 QAPAAAASAEFASATATAGASAPAAATPAASAPVAEASAPAAAPKAPVVAPOPPA 498
 QY 360 -EEETESGIFGYSYITLLAGGAALTALLLRLAQSK--RARRTTEESVPEEEDLDAA 416
 DB 499 EEESFSSLLGNPMALGLGLVALAGLVYRRQKPEQAQHQDSLLSQESTVMAGA 558
 QY 417 DD-----GLEITFAVETATPEPAPKNDVNDTLALDGSSEELSAP 458
 DB 559 NSLFGAAGQSDITSQHSYFGADFRIGGNGESNEVDPTAEADVITYAGRDVQAEELL--R 616
 QY 459 QTFDVEDTTPSNRIDLPDSLAAQNGILSGALTQDEETQKR---ADADWNAIESDTSVY 515
 DB 617 EALEQHPEQAILRLK-LMEIYANRQDA--HGQTTAEEMLAQVGRASPDM-----AEAAA 668
 QY 516 EPETENPNPNFVIVIDTPPEPESVAQPAENKPTVDTDFSDNLPSPNNHIOTEETASAKPAS 575
 DB 669 LGRKTPDNPFLYITVQ-----GDGHHEQVAAAD-----DRHGHAQ-----GAVAAA 708
 QY 576 PSGLAGFLKASGPETILEKTVAEVQT--PEELHDLFLKVYETDAVATAPETPDFN----- 628
 DB 709 GAALAGMGAAAAAAAEAFKFTVTGETTRRGEWTTVDPMGDFGMPSTKAPQLADLEPLES 768
 QY 629 ----AAADDLSALLOPAPAPSVEENITETVA-----ETP--DFNATADDLSALLOPSEVPA 678
 DB 769 FPAPAGEBITAPLOAE---VFQPLAEPEAPPALPQLPTDPAFHAGE---AFQPASVPP 822
 QY 679 VERNAAEIVADDLSALLOPAPAEPAVEENTVETVAETSDHEHTAADDLSALLOPAEVPAVEE 738
 DB 823 LHMDLSDL-SLDLNPAPVAETAV-----EPAPAAVAADSLPFAWQPADLP----- 868
 QY 739 NVTKTVAEI PDFNATADDLSALLOPSEVPAVEENAAEI-TLETPDGNSTSEADALPDFLKD 797
 DB 869 -----ETPVQLDAAPQFQEVVAQQBEPLTVMLDLTNLPHLSAEKGIDCVRD 915

QY 798 GBEETVMSIYLSEENIPNNADTSFSPSEVSGSDAPSEAKYDLAEWYLEIGRDAAAATVQ 857
 DB 916 -----LQIKFDLAKAYIEIGDKEGARELLQ 940
 QY 858 KLLEEARBGDVLKRAQALAAQELG 879
 DB 941 EVLIDLGDPSFHAEBAQALMRQIG 962
 RESULT 5
 Q87Y10 PRELIMINARY; PRT; 947 AA.
 ID Q87Y10
 AC Q87Y10;
 DT 01-JUN-2003 (TrEMBLrel. 24, Created)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Conserved hypothetical protein.
 GN PSTO3818.
 OS Pseudomonas syringae (pv. tomato).
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
 OC Pseudomonadaceae; Pseudomonas.
 EX NCBI_TaxID=323;
 EN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=DC3000;
 RA Buell R., Joardar V., Khouri H., Fedorova N., Tran B., Russell D.,
 RA Berry K., Utterback T., Van Aken S., Feldblyum T., Gwinn M.,
 RA Dodson R., DeBoy R., Durkin A., Kolonay J., Madupu R., Daugherty S.,
 RA Brinkac L., Beanan M., Haft D., Selengut J., Nelson W., Davidsen T.,
 RA White O., Fraser C., Collier A.;
 RA "Complete sequence of Pseudomonas syringae";
 RT Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.
 RL EMBL; AE016869; AA057286.1; -
 DR TIGR; PSPTO3818; -
 DR GO; GO:0016020; C:membrane; IEA.
 DR GO; GO:0005524; F:ATP binding; IEA.
 DR GO; GO:0004009; F:ATP-binding cassette (ABC) transporter acti. . .; IEA.
 DR GO; GO:0006810; P:transport; IEA.
 DR InterPro; IPR003439; ABC transporter.
 DR InterPro; IPR008941; TPR-like.
 DR PROSITE; PS00211; ABC TRANSPORTER 1; 1.
 KW Hypothetical protein, Complete proteome.
 SQ SEQUENCE 947 AA; 99429 MW; 84789769FCF9B69 CRC64;
 Query Match 8.8%; Score 387.5; DB 16; Length 947;
 Best Local Similarity 22.6%; Pred. No. 1.1e-10;
 Matches 230; Conservative 133; Mismatches 377; Indels 279; Gaps 39;
 QY 44 IAAVAVAAASFOAHAGLGLNLIQSLNDEPFSGSITVIGEEAKALLGGSVTVSEKGLTAK 103
 DB 11 IAAASALSSGMAQALGLGELSVKSTLNQPLVAEILT--DAQGL--NAAQVVPSLATTAD 66
 QY 104 VHKLK-----DKAVIAVSSEQVRDPVLVFRIGA---GAQVREY 140
 DB 67 FAQLGVSRQELNDLITFTFVIDPNGSVLRITSSKPVYVVKFLVQVLPNGRLVRDYS 126
 QY 141 AILDVPVGYSPKTKSALSDGKTHRKTAFTAESQ-----ENQNAKALRKTKDSANAAVKP 195
 DB 127 LLLDPPKYSPEAAAAAAPAPAPAAAPAAQATVAPSTEAAAAQAPAPAPAAATPAAPV 186
 QY 196 ---AYNGKTHTVRKGETVKQIAAATRPKHLTLEQVADALLKANPNVSAHG---RLRAGSV 249
 DB 187 APAADKQAQVYVANNITLWEIAAKVRTGG-TVQQTMLAQNLNPDFAFMGNINRLKKGQV 245
 QY 250 LHIPNLNRIKAEQPKQ-----TAKPAEATASMPSEPS 282
 DB 246 LRLPTQQTTA-LPQQAIAEVSQYSAWKEGRRLPTGTQVDAITRRDRAGAPSKVDTS 304
 QY 283 KQATVEKPVKEPKAK----- 297
 DB 305 DNLSLVSAAGGKPAAGKAGGAGDADLGNKLAVAQEAALDTRTRDNRAELKGRMDLQSLDKLQR 364

```

QY 298 -----VAAPEAKAEKPAVRPEPVPAAANTAASATAESAPQOE-- 333
Db 365 LIELKNGQLAKMQAAGAAVPPVAATPEATTPANAAVPAASLVANGAPVKPAGEIAPEDAL 424
QY 334 -AASALDTPDEGNNAVSEVQVSAEETESGLFGSYTLLIAGGGAALIALLLILRL 392
Db 425 PAGAQAQVATPAADQPLAV-BFVAATEDDDMLQKAL-DNPTILGLIGGAALLIALLLILFL 482
QY 393 AQSKRARR-----TEES--VPSEPDLDADAADGIEITTFVAETPA----- 431
Db 483 ARRAAKAEAEKHKMARALAESDFVDMMDAPQASFGDLVPPVNRVNRMGAAAGAA 542
QY 432 -----TPEPAKNDVNTLALDGESEELSAKQTFDVTETTPSNRIDLDFSLAAQ-NG 485
Db 543 AABERSADPLVQAEIHIAIYGRMNOQAVELLEBA-----VKDDPARDDIRLKMEIVAEQGN 598
QY 486 ILSGALTODEETQKADADNATSTDSVVEPEFNFYNEVVEIDTPEPESVAQTAENK 545
Db 599 KAYAHERKLUVAAGKEAEVEQLEERHS-----TLKEFVAPVVLFPVAPQASEPAPASAA 653
QY 546 PETVDTDFSNLPSNNHIGTEETASAKPASPSGLAGFLKASSPETILEKTVAEVTPEEL 605
Db 654 PAV-----AAAVAASAALAEALDAKYVEELLADDSAE-OPPEPV 693
QY 606 HDFLKY-----ETDAVAETAPETP-----DFNAADLISALLOPAEAPSVEENITETVAETP 658
Db 694 AEPEPVVPEPELAAAPVEEDPFDFDLSDLEFDEFTTTP--QVSTVDNLDLMLDEP 751
QY 659 DFNATADD--LSALLOPSEVPAVENAA--EIVAD--DLISALLOPAEAPAVE-- 705
Db 752 ALSAVPDEPLSFESIMQOE-----BARAATTPEDLADFDLDS-----EEDPALKNEDD 802
QY 706 -----NVTETVAETSDFTHTAADDLSALLOPAEVPVAVENNVKIVA-----EIPDFENATA 754
Db 803 FLILGSGEPLDGETKVPPISSDLE-LPEDFLSLADEIETDQASQAFATEIDVNA-- 859
QY 755 DLSALLOPSEVPAVENAAITLETDPNSNTSADALPDFLKDGEETVDMWSIYLSEENI 814
Db 860 -ELDRLAQNLEHPILDE-----PRPTAEDAAALDD----- 888
QY 815 PNNADTSPFSESVCSPASEAKYDLARMYLEIGDRDAAAEVTKLLEBAEGVILKRAQA 873
Db 889 --EPDFDF--MAGTD-EAATKLDLARAYIDMGADGARDILDEV--TEGDDGQKSEA 939

RESULT 6
QHZA6 PRELIMINARY; PRT; 919 AA.
AC Q9HZA6;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein PA3115.
GN PA3115.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 15692 / PA01;
RX MEDLINE=20437337; PubMed=10984043;
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,
RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA Garber R.L., Goltry L., Tolentino E., Westbrook-Wadman S., Yuan Y.,
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Sater M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an
RT opportunistic pathogen.";
RL Nature 406:959-964(2000).
DR EMBL; AB004735; AAG06503.1; -.
DR F01; F83257; F83257.

```

```

DR GO; GO:0016998; P:cell wall catabolism; IEA.
DR InterPro; IPR002482; LysM.
DR InterPro; IPR000408; Reg_chrom_condens.
DR Pfam; PF01476; LysM; 2.
DR SMART; SM00257; LysM; 1.
DR PROSITE; PS00626; RCC1 2; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 919 AA; 96928 MW; CE745BCFAD414A CRC64;

Query Match      8.7%; Score 381.5; DB 16; Length 919;
Best Local Similarity 23.0%; Pred. No. 2.1e-10;
Matches 232; Conservative 134; Mismatches 362; Indels 281; Gaps 43;

QY 41 IKLIASVAVAAAFQAH-A-GLGINTQSNLDEPFGSGITVTGEEKALLGGGVTVS--- 96
Db 8 VRAIAAA-SVLTSGMAHGLGLGITLKSALNQPLDAEIELL--EVRD-LGSEVIPSLAS 63
QY 97 -EKGLTAKVHKL-----GDKAVIASSEQAVRDVILVPRIGA---GAQVR 137
Db 64 PEFSKAGVDRLYLTLTKFTPVKPNKGSVIRTSKPVQEPYLNFLVQVLPNPGRLLR 123
QY 138 EYTAIILDPVGYSPKTSALSADGKTHRTAPTAEBSQENQNAKALRTDKKDSANAAYK-- 194
Db 124 EYTVLLDPPLYSPQAAASAPQAPV--SAPRATG-----APRAPQAPVETTA 169
QY 195 PAYNGKTHVTKETVKQIAAATRPKHLTLEQVADALLKANPNVSAHG---RLRAGSVLH 251
Db 170 PAGESDTRIV-SNDTILWEIAQRNTRDRVSPQAMLAFOELNPGAFVQDNTNRLKSGVLR 228
QY 252 IPNLNRIKAEQPKPQTKPKAETAS-----MPSEPSK-QAT 286
Db 229 IPTQQMLERSPREALSQVQAQNSQWRGNPAAGSAGARQLDATQRNAAGSAPSKVDAT 288
QY 287 ----- 286
Db 289 DNRLVSGEGKAGKADKGGKDSKAIADTLAVTKESLDSTRENEELQSRMQDLSQLD 348
QY 287 -VEKPVKEKVA-----APEAKAEKPAVRPEPVPAAANTA 321
Db 349 KLQKLIQLKDAQLAKLQGLGAGGQGAOPNAALPDASQPNAAQAQAPQGTPTAAAPT 408
QY 322 ASETAAESAPQAAASAIIDTPTDE--TGNVSEPVQVSAEETESGLFGSYTLLIAG-- 378
Db 409 APAGEPAAPQPPVAPPAPAEKPPAPVAPVAPVQAAEQAPSPFL-----DELANP 463
QY 379 -----GGAALIALLLLLLAQSKRARTTESVP-----EPEPDLDADAADGI-EITF 424
Db 464 LMLAVIGSALLALLVLLMILSRNAQKEKEBAQAFADTGEQEEDALDLGKDFDGLTL 523
QY 425 AEVETPATPEPAPKNDVNTLALDGESEEL-----SAKQTFDVTETTPSNRIDLD 475
Db 524 DEPE-PQVAAVAPQVEKTTAQTSDALGEADIIYAYGRFNQAAELLQNAIYDEP-QRTDLR 581
QY 476 FDSLAA-AQNGILSGALTQDEETQKADADNATSTDSVVEPEFNFYNEVVEIDTPE 534
Db 582 LKLMEVYAEIMGDREGFAQENELRIGGQ-QPVQLKSRV-----PAMVAVAAVA 631
QY 535 PESVAQTAENKPEVTDTDFSNLPSNNHIGTEETASAKP-ASPSGLIAGFLKASSPETILE 593
Db 632 GLAGAKLAQDELDSPSLD-DLSLDDSGH-----AAKPDAAQDLDLDAFDSLDDLGDD 683
QY 594 KTVASVQVPE-ELHDFLKYETDVAETAPETPDNNAADDLSALLOPAEAPSVEENITE 652
Db 684 DVQADLKSDSGALDLDLTLSDLDLAASTPADKP-----VDDLDFGLDFAE----- 728
QY 653 TVAETPDFNATADDLSALLOPSEVPAVENAAEIVADDLSALLO--PAEAPAVENVTE 709
Db 729 -LAETPS-OPKHDDLDGDFSLDLADP--EDKUSD---DDFLISLNDENVFAAPAEFTLD 781
QY 710 TVAETSDFTAAD--DLSALLOPAEVPVAVENNVTKVAEIPDFNATADDLSALLOPSEVP 767
Db 782 TEAAEPEALSLPDDFDLSLADEPTPEAPAEKGEDSFAAQLDVSAQLDELAS----- 833

```

```

QY 768 AVENAAEITLETDSNT---SEADALPDFLKOGEEETVDSIYLSEENIPNNADTSPFS 824
Db 834 -----NLDEPKSATPFSADAASALDGD-----ADDDDF 866
QY 825 ESVGSDAPSEAKYDLAEYMLIGDRDAAAETVQKLLBEAEGVILKRAQA 873
Db 867 LSGADERAT--KLRLARAYIDMGSEGDARDILDEVL--AEGNDGQAQA 911

RESULT 7
Q9KTA5 PRELIMINARY; PRT; 1621 AA.
ID Q9KTA5
AC Q9KTA5;
DT 01-OCT-2000 (TremBLrel. 15, Created)
DT 01-OCT-2000 (TremBLrel. 15, Last sequence update)
DE 01-JUN-2003 (TremBLrel. 24, Last annotation update)
DE Hypothetical protein VC0998.
GN VC0998.
OS Vibrio cholerae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC Vibrionaceae; Vibrio.
OX NCBI_TaxID=666;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=E1 Tor N16961 / Serotype O1;
RX MEDLINE=20406833; PubMed=10952301;
RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,
RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
RA Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,
RA McDonald L., Utterback T., Fleischmann R.D., Nierman W.C., White O.,
RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
RA Fraser C.M.;
RA "DNA sequence of both chromosomes of the cholera pathogen Vibrio
RT cholerae.";
RL Nature 406:477-483(2000).
DR EMBL; AE004181; AAF94159.1; -.
DR PIR; A82255; A82255.
DR TIGR; VC0998; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 1621 AA; 178073 MW; 2EF38BAC27A2485D CRC64;

Query Match 8.0%; Score 353.5; DB 16; Length 1621;
Best Local Similarity 23.1%; Pred. No. 1e-08;
Matches 226; Conservative 129; Mismatches 316; Indels 307; Gaps 50;

QY 137 REYTAILDPVGSPTKMSLSGKTHRTAPTAESEQENQAKALRKTOKDSANAARPA 196
Db 24 RFFQRLLPVAVMVVVTQTSFVSAESTRLVGPQGVQPTPO---YSENIVRNSANN--EPG 78
QY 197 -YNGKTHTVRKGETVKQIAAARIP-KHLTLEQVADALLKANP---NVSAGRLRAGSVL 250
Db 79 RFFGPTSA---NQTLWSIASQLRPSSTVQTLAIYQINQAPENQNIH-TLIFGSTL 134
QY 251 HIPNLNRKAEQPKQPTAKPKPAETASMPSEPKQATVEPKPE-AKVAAP--EAKAEK 307
Db 135 RVPSLAQISNSTQDAVNIMASHQAKINQTPD---TPVRPVAPPAPVATPKVAVQAT 191
QY 308 PAVRPPVPAANTAASETAASAPQAAASADTPT----- 343
Db 192 P---POVTP-----TAQPKAPTCLKTPAKPSQSTDAEYMALEEKNHTRLML 237
QY 344 -----DETG--NAVSEPEVQVSAE-----ETESGLFGSYTLLAGGG----- 380
Db 238 SQVQSEVSTLKEELGDNIRSEVERLLEERKAEASRLAPSLDNLNLSNGWLVALLA 297
QY 381 -----AALIALLLLRQAQSKRARTESVPPEEPDLDAADDGIEITFAEVETPATPE 434
Db 298 LIPGLLJIAIVLLNRRSSAQENPTNNITSEMP-----TAAPVTLG 341
QY 435 PAPKNDVNTLALDG-----ESEELSAQTF-----DV-----ETDTPSN----- 470

```

```

Db 342 PEQTEIDGDDLLDDDLFSTTDDKEENDAERKAFSDEDDVPADLNELTDLDFNLGQSDDL 401
QY 471 -----RIDLPDLSLAAQNGILSGALQTDE-----ETQKRA-----DADNVAI 508
Db 402 FVGIDDDGDLDTFEDALNESANGI---SVNADKALGLEEMERALNDVSEPTDNDLNSF 458
QY 509 ESTSVVPEPTFNYPNVEIVIDTPEPE-----SVAQTAENKP----- 546
Db 459 DLADENQWSE-----DDIEALLSGDERNELLSGKVQSLDLDLLASELDALDDPEALQD 513
QY 547 -ETVDTDFSNLPSNNHIGTEETASAKPASPSGLAG-----FLKA 585
Db 514 TETLDTLLNDELASL-----SEDDDEFDLSGAGVAGDQDLDDLPASIEEQADLEQLEAKA 569
QY 586 SSPETILEKTVAEVQTP-----EELHDFLKVYETDAVETA-----PETPDFNAAA 631
Db 570 IDETALLDEILAEQDAPLSEESTELLDELLDDFKPENDEFAQTADLLOPEEPILDEE 629
QY 632 DDLSALIQ-----PAEAPS---VEENITETV-----AETPDFNATADDLSA 669
Db 630 DSTQLLNEVLGEVPPEELASGLEIDQNSTELLDELLDDLDDESIEATEFSVAPEKLSV 689
QY 670 -----LLQPSVEPVAEENAAEIVADD-----LSALLQPAEA-----PAVENV 707
Db 690 EDGTELFDELLEIEQHPESASLPELATEDEFNSDTFIDLLNGAPAKDPLLEPVLDENE 749
QY 708 TETVAETSDPHT-----AADLSALLQPAEVA-----VEENTKTV---AE 746
Db 750 APAQADDFPNFETEGGLEDDL---QPSALPANFETGPDDEWVDFDDSSPTLEGNAE 806
QY 747 IPDFNATADDLSALLQPSVEPVAEENAAEITLE---TPDSNTSEA---DAL-PDFLKDGE 800
Db 807 L-ELSSAEDDL-----PEQTATNETADELLADLAAQPSQNTVTSDDALAPDGLSQSV 860
QY 801 ETVDWS-IYLSEEN-IPNNAD---TSPPSSVSGSDAPSEAKYDLAEYMLIGDRDAAA 854
Db 861 EPLTLDLELPEENDEPQLAEVTPSSAFDEQVETEIRPESEPLAAEASNDSDLTALNE 920
QY 855 TVQKLLBEAEGDVLKRAQ 872
Db 921 L--DLPEYTEEDVLADVQ 936

```

RESULT 8

```

Q8DB34 PRELIMINARY; PRT; 1951 AA.
AC Q8DB34;
DT 01-MAR-2003 (TremBLrel. 23, Created)
DT 01-MAR-2003 (TremBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TremBLrel. 25, Last annotation update)
DE AAA ATPase.
GN VV11991.
OS Vibrio vulnificus.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC Vibrionaceae; Vibrio.
OX NCBI_TaxID=672;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CMCP6;
RA Rhee J.H., Kim S.Y., Chung S.S., Kim J.J., Moon Y.H., Jeong H.,
RA Choy H.E.;
RT "Complete genome sequence of Vibrio vulnificus CMCP6.";
RL Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE016803; AAO10389.1; -.
DR InterPro; IPR001440; TPR.
DR InterPro; IPR008941; TPR-like.
KW Complete proteome.
SQ SEQUENCE 1951 AA; 212759 MW; 73DBF9E1EED8DEE7 CRC64;

```

Query Match

Best Local Similarity 22.6%; Pred. No. 2.7e-08;
Matches 185; Conservative 137; Mismatches 305; Indels 192; Gaps 36;

SQ SEQUENCE 2768 AA; 294032 MW; CA929A21774E4684 CRC64;

Query Match
Best Local Similarity 7.3%; Score 319.5; DB 5; Length 2768;
Matches 198; Conservative 139; Mismatches 323; Indels 321; Gaps 40;

QY 101 TAKVHKLGDKAVIAVSQAVRDPVLVPRRI-----GAGAQRVRE-YTA 141
DB 1141 TAKPDNKIDVSEISIEIEIPKDVIMPTGIEQPLSHVQPDERTQVTSVPAQDESITTA 1200
QY 142 ILD--PVGYSPKTSALSADGKTHRTAPTASQEN-----QNAKALRTYDKDSANA 191
DB 1201 KVDKKPIDESAEDKKPI--GESEDSKPIDSESDKKPVEESAEDKKPVEDSEKKEPLP 1258
QY 192 AVKPAYNGKTHVRKEIVKQIAAAI-----RPXHLTLEQVADALLK-----ANPNVSA 240
DB 1259 TVIIPA--SEIEKESPEDEKTEADFAAPTQPEATTQAQADTAKEVEDDKLATTSPV 1316
QY 241 HGRLRAGSVLHPLNLRKAEQPKQTAQKAEATASMPSEPSQATVKEPKVEKPAKVAA 300
DB 1317 SGE-----DELKPADEKKTETAQIP-----DAEIPASTDEPESSTEL 1354
QY 301 PEAKAEK-----PAVRPEVPVPAANTAAAS-ETAABE-----APQEAASAIITDP 342
DB 1355 PTVLDLKKPEEDSTKGTAPESDKVPEVPTASTENEIEESDKFTTVAPPKISADETEP 1414
QY 343 TDETG--NAVSEPVQO-----VSAAEETESGLFGGSY 372
DB 1415 TAEDLVATPEFIESEFVSTKPAVQGPPLPTLAPQPEKKPVDATSTHEADI-----1469
QY 373 TILLAGGGAALIALLLLRLAQSKRAARTEESVPEEPDLDDAADDG-----IETFAEV 427
DB 1470 -----STEPSAEVEKEASGETSESDNEIDAGASSTVPVPSADED 1508
QY 428 ETAPTEPAPKNDVNDTLA-LDGESEBELSAQTFDV-ETDTPSNRILDFDLSLAAQNG 485
DB 1509 KTPSTKTEVADDKFTTVAPLAGDEESNLPLQDIFEEAPV-----A 1553
QY 486 ILSGALTODEETQKRAADNNALIESDTSVYBETFPNPNV-----524
DB 1554 VTTPAAPSDDGQKQVPEVEKEPIEDGQKPIEDTSTPTSSSENEIPESDRATTIAPSKEE 1613
QY 525 -----PVEIVIDTPEPSVAQTAENKPTVDVTDPSDNLPSNNHIGTTEE-TAS 570
DB 1614 PSEPSTGAPTQKDEPAPSTDAFESDRSKETPESEVPTTVAPAGEKIPTSSITPOSEPTAT 1673
QY 571 AKPASPSGLAGFLKASSPTILEKTVAEVQTPPEELHDFLKVYETDAVATAPETPDFNAA 630
DB 1674 SAP-----VAKPDEDEK-----ETSTEI-----PTDAPASS-----EEDENS 1707
QY 631 ADDLSALLQAPAEAPSVEENITETVAETPDFNATADDLSALLQSEVPVAVEENAAETVADD 690
DB 1708 TQOI-----PSEVP-----EKKPTPAQTREGDIVGTAAPTTSDEVPVQRLPEVLAIE- 1758
QY 691 LSALLQAPAPAVEENVTVAETSD-----PHTAADLDSALLQOP-----730
DB 1759 ---IPQSTETGKQDQETTAAPSIDRKEPVVTEIDEAETTVAPISEKDEKPTEEKPVYE 1815
QY 731 -----AEVPVAVENVTKV-----AEIPDFNATADDLSALLQSEV-PAVEENAAETITET 780
DB 1816 OKPTGEPSSEEEKEKPIEQDVSTEGP-VSTEASEAGSTESSESEVKEPSTEGEVAEKPEDK 1874
QY 781 PDSNTSEA--DALPDF-----LKDGEEETVDMISIVLSEENIPNNAOTSPPSE-----825
DB 1875 QPSSTAQAQVETIPEISTELPAQDGKPT-----SEAPVDSDEDTSPSDEKIPSVSG 1927
QY 826 -----SVGSDAPSEAKYDLAEM-----YLETIGDRDAAAEATVQKLLLEEA 863
DB 1928 EEVEGPEVTTASPOAAEEDELKTPAESSEPSSTDKVPETEXQKPEDEKADETPESVTQVS 1987
QY 864 E-----GDVLKRAQA 873
DB 1988 DVATSTAPVAGGDIEKDEQA 2008

RESULT 13
Q07290 PRELIMINARY; PRT; 1822 AA.
ID Q07290
AC Q07290;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
DE Epf* protein.
GN Epf*
OS Streptococcus suis.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1307;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1890;
RX MEDLINE=93328288; PubMed=8335363;
RA Smith H.E., Reek F.H., Vecht U., Gielkens A.L.J., Smits M.A.;
RT "Repeats in an extracellular protein of weakly pathogenic strains of
RT Streptococcus suis type 2 are absent in pathogenic strains.";
RL Infect. Immun. 61:3318-3326(1993).
DR EMBL; X71880; CAA50714.1;...
DR PIR; S33441; S33441.
DR GO; GO:0016020; C:membrane; IEA.
DR InterPro; IPR005877; Gpos_YsIRK.
DR Pfam; PF04650; YsIRK_signal; 1.
DR TIGRFAMs; TIGR01168; YsIRK_signal; 1.
DR SEQUENCE 1822 AA; 192632 MW; 3838960C77641D7D CRC64;
Query Match 7.2%; Score 316; DB 2; Length 1822;
Best Local Similarity 19.9%; Pred. No. 7.8e-07;
Matches 204; Conservative 165; Mismatches 363; Indels 294; Gaps 43;

QY 37 NNRQIKLTAASV-----AVAAAPQAHAGLGLNIQSNLD--EPFSGSITVTGEAKAL 87
DB 682 DNERLKLGLPDSFTVNSDGTVSVDYSA-----GGVNDGATDIKNAATNLATRNQAKAE 737
QY 88 LOGGSGVTVEKGLTAK-----VHKLGDKAVIAVSQAVRDPVLVFRIGAGAQRVREYTAIL 143
DB 738 I-DTKLAEHKKAEIAKRDFAFSKIDDISLRAEQQAADKAAVAA--AAGDALKE-----L 789
QY 144 DPVGVSPK-----TKGALSDGKTHRTAPTAESEQENQAKALRTYDKDSANAAY--- 193
DB 790 DNKATEAKEKIDKATTASEINDAKTNGEI--NLDSEAEVGEKAIINQAKAEKELAKAEVKNK 847
QY 194 -----KPAY-----NGKTHVRK--GETVQKIAAAIRPKHLT-----223
DB 848 APEALEKVNNNPDLLEBKAYPDDIKESKEVAVKINNAENTAETITAEIAENED 907
QY 224 -----LEQVAD---ALLKANPNVSAHGRLRAGSVLHPLNLRKAEQPKQTA 268
DB 908 VINAAQLDALNKLKXSEETKAAIDANPNLTPEEKAKA-----IAKVEELVNAESDILS 962
QY 269 KPAETASMPSPSQATVKEPKVEKPAKVAPEAK---AEKPAVRPEVPVPAANTAASET 325
DB 963 KPTPTVQAVEDKA-----DKDLAKVELQAAADGAKKGEANPNTPEEKVAKKAVEADA 1017
QY 326 AAEAPQEAASAIID--TPTDETGNVAVSEPVQVSAEE--ETESGLFGSVTLILLAGGG 380
DB 1018 V-----KVATDAIDKASTPT-EVDTATSDGVKAIADAEFEKATOKD-----1056
QY 381 AALIALLLLRLAQSKRAARTEES-----VPEEPDLDDAADDGIEITFAEVETP 430
DB 1057 -----AKNKIAKEAESAKKAIIDNPNLTPEKESAKNAVEEAARVATAIDKA 1104
QY 431 ATPEPAPKNDVNDTLADGSESEELSAKQTFDVEITTP-----SNRIDLDFDLSLAAQ 483
DB 1105 STPD-----AVQVEEDKGVAAILNITAKADAKGVIAKLADEIKLEKDKQAE 1153
QY 484 NGILSGALTODEET-----OKRADADNNAIESTDSVYBETFPNPNVPEI-----528


```

Db 1154 KADASTWNEKAKAIALQDVVDKGALEDAARVATNEIIBATTTEKAKAEALAGEK 1213
QY 529 -----VIDTPBPESVAQTAENKPTVDTFDSNLPNNHIGTEETASAKPASPS 577
Db 1214 SLTDTGKEARDAVELAKDELAKELAEIRTEEBEATKIVEKLAEDTRKAIENPNLSDEKQ 1273
QY 578 GLAGFLKASSPFTI-----LEKTVAEVQTPPEELHDFLKVYETDVAETAPETPDFNAAA 631
Db 1274 AEIKKLTDAVAKTLATIRDNADKRTQBAEKQAQALADLEKAKETQKIAD-----KAAI 1325
QY 632 DLSALLQPAEAPSVEEN-----LITEVATPFDEN-----ATAD 665
Db 1326 DRILTILVKGELBATQDKAKKIADKAAAKEAIAFNPLTDAEKKTTFTDAVDAEVAKAN 1385
QY 666 D-LSALLQPSVEPAVENNAEIVADD-LSALLQPA-----EAPAVE-----NTE-- 709
Db 1386 DAISAATSPADVQKEEDAGVAAAEVLDVRAKQDKAKKIADKAAAKEAIGSPNLTDAE 1445
QY 710 -----TVAETSDFHTAADDLSALLQPAEVPVAVENVTKTVAEIPDFNATADDLSALLQPS 765
Db 1446 KKTFTDAVDAEAVAKANDAIASAATSPADVQKEEDAGVAAIAE-----DVLDAAKQDAK 1497
QY 766 VPAVENNAEITLETDPDNTSEADALPDFLKDGEEETVDWSIYLSSEENIPNADTSPSE 825
Db 1498 NTKAKES-----DAKSAIDAMPN-LTDAEKESAKKAV-----DADAKAATD 1538
QY 826 SV-GSDAPSEAKY-----DLAEMYLEIGDRDAAAEATVOKL-----LEAEAG 865
Db 1539 AIDASTSPVEAQSAEDKGVGSIADQVLDRAKQDKAKKIADKAAAKEAIDANFNLSDAEK 1598
QY 866 DVLKRA 871
Db 1599 EASKKA 1604

RESULT 14
Q9U0P0 PRELIMINARY; PRT; 1786 AA.
AC Q9U0P0;
DT 01-MAY-2000 (TremBLrel. 13, Created)
DT 01-MAY-2000 (TremBLrel. 13, Last sequence update)
DE 01-OCT-2003 (TremBLrel. 25, Last annotation update)
DE Liver stage antigen-3 precursor.
GN LSA-3.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K1;
RA Daubersies P., Thomas A.W., Millet P., Brahimi-Zeghidour K.;
RT "LSA-3, a conserved pre-erythrocytic malaria antigen can induce
RL protection in chimpanzees";
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ007010; CAB65343.1; --
DR GO; GO:0003723; F:RNA binding; IEA.
DR InterPro; IPR001313; Pumilio/Puf.
KW Signal.
FT SIGNAL 1 63 POTENTIAL.
FT CHAIN 64 1786 POTENTIAL.
SQ SEQUENCE 1786 AA; 200101 MW; 5DF536D7B5B1BD98 CRC64;

Query Match
Best Local Similarity 20.2%; Pred. No. 8.1e-07;
Matches 167; Conservative 147; Mismatches 331; Indels 183; Gaps 26;

QY 64 NIOSLDEFFSGSITVTGEEAKALLGGSVTVSEKGLTAKVHKLGDKAV---IAYSSEA 120
Db 242 NVEENVNEDDGSVASSVEES-----IASSVDESIDSSIEENVAFTVEEI 286
QY 121 VRDPVLVFRIGAGAVREYTAILDPVGYSPKTSALSODGKTRKTPATAES---QENQAK 178
Db 287 VAPSVV-----ESVAPSVEESVEENVEES 310

```

```

QY 179 ALRKTCKDSANAAVKPAYNGKTHTVRKGE-----TVKQIAAAAIRPKHLTLEQ-VADALL 232
Db 311 VAENVEESVAENVEESVAENVEESVAENVEEIVAPTVEEIVAP-----TVEEIVAPSVV 364
QY 233 KA-NPNVSAHGRLRAGSVLHPNLNRIKAPQPKQTAQKPAETASMPSESPSKQATVEKPV 291
Db 365 ESVAPSVEESVE-----ENVEESVAENVEESVAENVEESVAENVEESVAENVEESV 415
QY 292 EKPEAKVAPE-----AKAEKPAVRPEPVPAAATAAETAESAPOQAAAAIDTPTDET 346
Db 416 AENVEEIVAPTVEEIVAPTVEEIVAPSVVSVAPSVEESVEENVEESVAENVEESVAENV 475
QY 347 GNAVSPVEQVSAEETESGLFGSGTYLLLAGGGAIALILLRLAQAQRARTTESVP 406
Db 476 EESVAENVEESVAENVEES-----VAENVEESVA 504
QY 407 EE-EPDLDDAADGCIITTAETPATPBAPKPNVDNLTALDGESEELSAKQTPDET 465
Db 505 ENVEESVAENVEEIVAPTVEEIVAPTVEEIVAPSVVSVAPSVEESVEE-----NVEE 557
QY 466 DTPSNRIDLDFSLAAQNGILSGALTQDEETOKRADAMNAIESTDSVYEPETFPNYP 525
Db 558 SVAEN-----VEESVAENVEESVAENVEESVAENVEEIVAPTVEEIVAP 601
QY 526 -VEIVIDTPPESVAQTAENK-PEYVDTDFSDNLPNNHIGTEETASAKPASGLAGFL 583
Db 602 TVEEIVAPSVVSVAPSVEESVEENVEESVAENVEESVAENVEESV----- 647
QY 584 KASSPETILEKTVAEQTPPEELHDFLKVYETDAVETAPETPDFNAAADDLSALLQPAEA 643
Db 648 -AENVEEIVAPTVEEIVAP-TVEEIVAPSVVSVAPSVEESVEEN-VEESVAENVEESVA 704
QY 644 PSVEENITETVATPDFNATADDLSALLQPSVEPAVEENAAEIVADDLSALLQPAEAPAV 703
Db 705 ENVEESVAENVEE-----SVAENVEEIVAPTVEEIVAPTVEEIVAPSVVSVAPSVEESV 759
QY 704 EENVTETVATSDPHTAADDLSALLQPAEVPVAVENVTKTVAEIPDFNATADDLSALLQ 763
Db 760 EENVEESVAENVE-----ESVAENVEESVAENVEESVAFTVEEI-----VAPSVEE 805
QY 764 SEVPVAVENNAEITLETDPDNTSEADALPDFLKDGEEETVDWSIYLSSEENIPNADTSP 823
Db 806 SVAPSVEESVA-----ENVATNLS-D-NLLSNLGGTETEIKDSILNEIEEVENVVTI- 859
QY 824 SESVGSADPSEAKYDLAEMYLEIGDRDAAAEATVOKLLEAEAGDVLKRA 871
Db 860 LENV-EETTAESVTTFSNILEBIQENTITNDTIEEKLLEHENVLSAA 906

RESULT 15
Q8XTC4 PRELIMINARY; PRT; 673 AA.
AC Q8XTC4;
DT 01-MAR-2002 (TremBLrel. 20, Created)
DT 01-MAR-2002 (TremBLrel. 20, Last sequence update)
DT 01-JUN-2003 (TremBLrel. 24, Last annotation update)
DE Probable transmembrane protein.
GN RSP0189 OR RS04701.
OS Ralstonia solanacearum (pseudomonas solanacearum).
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Burkholderiaceae; Ralstonia.
OX NCBI_TaxID=305;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GM11000;
RX MEDLINE=21681879; PubMed=11823852;
RA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,
RA Ariat M., Billault A., Brottier P., Camus J.C., Cattolico L.,
RA Chandler M., Choisne N., Claudel-Renard C., Cunnac S., Demange N.,
RA Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schiex T.,
RA Siguier P., Thebault P., Whalen M., Wincker P., Levy M.,

```

Search completed: September 2, 2004, 19:20:43
Job time : 156 secs